Avoiding Ecological Fallacy: Assessing School and Teacher Effectiveness Using HLM and TIMSS Data from British Columbia and Ontario

By

Yichun Wei

B. Eng. University of Hunan

M. Sc. University of Hunan

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfillment of the requirements of the degree of

DOCTOR OF PHILOSOPHY

Faculty of Education

University of Manitoba

Winnipeg, MB. Canada

ABSTRACT

There are two serious methodological problems in the research literature on school effectiveness, the ecological problem in the analysis of aggregate data and the problem of not controlling for important confounding variables. This dissertation corrects these errors by using multilevel modeling procedures, specifically Hierarchical Linear Modeling (HLM), and the Canadian Trends in International Mathematics and Science Study (TIMSS) 2007 data, to evaluate the effect of school variables on the students' academic achievement when a number of theoretically-relevant student variables have been controlled. In this study, I demonstrate that an aggregate analysis gives the most biased results of the schools' impact on the students' academic achievement. I also show that a disaggretate analysis gives better results, but HLM gives the most accurate estimates using this nested data set.

Using HLM, I show that the physical resources of schools, which have been evaluated by school principals and classroom teachers, actually have no positive impact on the students' academic achievement. The results imply that the physical resources are important, but an excessive improvement in the physical conditions of schools is unlikely to improve the students' achievement. Most of the findings in this study are consistent with the best research literature. I conclude the dissertation by suggesting that aggregate analysis should not be used to infer relationships for individual students. Rather, multilevel analysis should be used whenever possible.

Keywords: HLM, multilevel, aggregate, ecological fallacy, school, TIMSS, mathematics

ACKNOWLEDGEMENTS

Near the end of this most challenging and rewarding endeavor, I want to thank all of the people involved in making this degree a reality. Foremost, I want to express my gratitude to the chair of my committee, Dr. Rodney Clifton, who guided me with such grace and dignity. Dr. Clifton, you have instilled confidence in me ever since our very first meeting. You also gave me the encouragement and gentle pushes I needed to keep me going when my spirit was down. Your demeanor had the calming and encouraging effect that I needed as I journeyed through this process. Thank you for sharing your knowledge and wisdom with me.

I also want to thank Dr. Robert Renaud, Dr. Lance Roberts, and Dr. Colleen Metge for their input and guidance. The members of my advisory committee have been very patient and they have been dedicated to insuring that I complete this degree. I also want to thank Dr. Xin Ma, my external examiner, for his valuable suggestions. Special thanks go to Dr. Romulo Magsino, who was my advisor for the first year in the program. Finally, I want to thank Julianna Enns in the Faculty of Education for her patience with my many questions.

DEDICATION

This work is dedicated to my husband, Shuang Hu, who has given continuous support and encouragement throughout the entire process; to my parents, Lanwen Shang and Xin Wei, who have always supported my academic endeavors; and to my two sons, Garry and Owen, who were born during my PhD program and to whom I will hopefully instill a love of learning.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

It is evident that some schools do significantly better than others in nurturing their students' academic achievement after considering the students' backgrounds. But, which factors contribute to the higher success of some schools, and especially the degree that they contribute to the students' academic achievement, is still an unresolved research question. There are, in fact, many contradictory and loosely defined conclusions in this research literature, among which, the effect of the schools' physical resources on the students' academic achievement has become a highly contested concern. A growing body of research has indicated the positive effect of the schools' physical resources by linking the qualities of school buildings that are related to the physical comfort of students to their attitudes and academic achievement (Berner, 1993; Buckley, Schneider, & Shang, 2004; Cash, 1993; Earthaman, 2004; Lewis, 2000; Roberts, 2009; Tanner, 2000; Tanner & Lackney, 2006; Uline & Tschannen-Moran, 2008; Uline, Tschannen-Moran, & Wolsey, 2009). Nevertheless, other researchers suggest that there is very little empirical support for the belief that the quality of the schools' facilities is a positive factor in the students' academic achievement (Hanushek, 1997, 2003; Hanushek, Rivkin, & Taylor, 1996; Henry, Fortner, & Thompson, 2010; Picus, Marion, Calvo & Glenn, 2005). It is difficult to determine any definite and consistent findings. In many cases, the researchers could not explain why the results obviously did not support the common sense point that the better physically resourced schools increase the achievement of students. For example, Cash (1993) found that higher rated schools unexpectedly had more disciplinary problems than lower-rated schools. Similarly, Earthman, Cash, and Van Berkum (1995) found that when considering the structural factors alone, students in substandard buildings outscored those in above-standard buildings.

Of course, the inconsistencies in this research literature may result from the different ways that the quality of physical resources has been measured. Roberts (2009) noted that the relationship between the condition of school buildings and the achievement of students is not evident if the school facilities are measured from the engineering perspective (e.g., FCI—Facility Condition Index) because it takes no account of the educational purpose of the schools. In contrast, measures based on the principals' perceptions of the educational requirements of the schools (e.g., CAPE—the Commonwealth Assessment of Physical Environment), tend to show a positive relationship with the learning outcomes of the students. No matter how the physical resources are measured, many studies suffer from limited variance in the measure of the schools' physical resources because they have focused on measuring the tangible aspects of the physical resources (e.g., quality of air, light, spatial density, etc.) and not the intangible aspects (maintenance, management, fitness, school culture, and virtual environment) or how the physical resources are used (Blackmore, Bateman, Loughlin, O'Mara, et al., 2011). It is likely that the physical resources of most schools are adequate for their basic educational functions in developed countries such as the United States and especially Canada. Consequently, the limited variance in the independent variable means that it shows little, if any, effect on the dependent variable. Thus, how the physical resources in schools are used is probably more important than the physical resources that they actually have.

Nevertheless, it is possible that these inconsistent findings result from two significant analytical problems. The first problem is the aggregate analysis commonly used in studies of the physical resources, in which the teacher and student variables are aggregated to the school level and then the school is used as the unit of analysis. For statistical reasons, the coefficients in aggregate analysis are most often overestimated (see Robinson, 1950). Moreover, it is a serious bias to assume that the relationship among variables at the school level would be the same as the relationship among the same variables at the student level. Assuming that these coefficients at different levels are identical is often referred as an ecological fallacy. The second problem is the lack of controlling for important confounding variables, such as the students' demographic backgrounds, their motivation to study, their teachers' effectiveness, or the social climate in their schools. Unfortunately, these variables are rarely available in data sets or these data sets are not large enough to use improved statistical methods, such as multilevel modeling, to control for confounding variables. In aggregate analysis, even if these variables are available, it is difficult, if not impossible, to control for their confounding effects. Without controlling for the confounding variables, it is impossible to draw definitive conclusions about the effects of schools on the students' academic achievement.

The Problem

Given these two methodology problems in the research literature on school effectiveness, this study used multilevel modeling procedures, specifically Hierarchical Linear Modeling (HLM), and the Canadian Trends in International Mathematics and Science Study (TIMSS) 2007 data, to evaluate the effect of school variables on the students' academic achievement when a number of student variables were controlled. More specifically, the impact of four school variables (the school's physical resources, instructional resources, average socioeconomic status or SES of students, and administration) and eight teacher and classroom variables (the teacher's gender, experience, instructional time, assignment of homework, attitudes towards teaching, the homogeneity of students in classroom, and the physical and instructional resources available in classroom) that have generally been assumed to affect the students' achievement were assessed. Within a two level model, these school, teacher, and classroom variables were analyzed at the higher level. Because students respond differentially to the learning opportunities they experience, five student variables (the student's gender, SES, instrumental motivation, educational expectations, and effort in completing homework) were examined at the lower level. At the same time, the provincial differences between British Columbia and Ontario were evaluated by analyzing each province separately.

TIMSS is an international program designed to assess the performances of students in mathematics and science at the fourth and eighth grades. As such, students write standardized mathematics and science tests. The Canadian TIMSS 2007 data set used in this dissertation provided a large nested sample of schools and students, and both the schools and students were weighted to represent the population of the fourth and eighth grade students in the two provinces. Specifically, 4,256 students were nested within 150 schools in British Columbia and 3,448 students were nested within 176 schools in Ontario. In order for the researchers to obtain information on the home and school environments of the students, the students, their mathematics and science teachers, and their school principals were asked to complete different questionnaires. Essentially, the Canadian TIMSS 2007 data is appropriate to assess the research problem in this dissertation. With improved modeling procedures, this study helps clarify some of the inconsistencies in the published research literature on school effectiveness, especially the biases caused by the aggregation of data.

Significance

Multilevel modeling could be used in school effectiveness research simply because schooling is multilevel. In the Canadian educational system, students are grouped within classrooms or by teachers; classrooms or teachers are grouped within schools; and schools are grouped within districts; which, in turn, are grouped within provinces. Briefly, each higher level provides each of the lower levels with contextual conditions that affect the students' academic achievement. Naturally, school effectiveness is a cross-level effect because the educational influences on students occur in the groups to which the students belong. Through both direct and indirect cross-level effects, the students' academic achievement is affected.

This multilevel view of schooling actually implies that the school characteristics cannot be conceived as being treatments that are uniformly administered to all students within a specific school, nor can they be conceived as being a constant to the development of each student in this school. Instead, the resources for instruction and therefore, the students' opportunities for learning, vary across schools, classrooms, and students. In addition, given the same opportunities for learning, students respond quite differently. Naturally, an adequate conceptualization of the schooling effects must include not only how schools differentially allocate resources and create learning opportunities for students, but also how students differentially respond to the available opportunities. As such, the multilevel framework of schooling must be used so that the cross-level interaction effects between certain conditions provided by classrooms/teachers, schools, (and even higher administration levels because they have some power over classroom and school levels) and students are all taken into consideration in the analytical procedures.

While this view of schooling is not difficult to understand, how to model the cross-level effectiveness of schooling is not obvious. Often, this cross-level perspective is modeled in two ways, disaggregate and aggregate analyses, both of which evaluate the cross-level impact of variables at a single level. The former distributes the data from the higher level units, the schools, to the lower level units, and thus these data become, for the purpose of analysis, attributes of the lower level units, the students. Then the lower level units, the students, are used as the unit of analysis. The latter procedure aggregates the data at the lower level units, the students, to the higher level units, the schools, and uses the higher level units as the unit of analysis. Unfortunately, both procedures are flawed. The disaggregate analysis is inappropriate because it ignores several basic assumptions of parametric analyses, specifically, the normal distribution of data, the homogeneity of variance, and the independence of observations. The standard errors calculated in the analysis are misleadingly small, and consequently, the null hypothesis is rejected more often than expected. A more significant problem, however, exists when the data are aggregated because researchers often assume that the relationships at the group level are the same as they are at the individual level, which is usually not true. Ignoring the disparity between ecological-level and individual-level relationships often results in researchers stating an ecological fallacy.

Conceptually, an ecological fallacy refers to the incorrect assumption that the relationships between variables observed at the aggregate, or ecological, level are necessarily the same as the relationships between the same variables at the individual level. This implies that the pattern of correlations displayed at the higher, or ecological, level cannot be simply generalized to individuals. Normally, ecological coefficients are much higher than individual-level coefficients. More importantly, ecological coefficients are often not realistic because unlike an analysis at the individual level, an ecological analysis cannot link an individual's outcome to the individual's exposure to the context variables, a problem which makes it almost impossible to infer the relationships at the individual level from ecological coefficients (Greenland, 2001; Greenland & Morgenstern, 1989; Greenland & Robins, 1994; Morgenstern & Thomas, 1993).

Unfortunately, many people in both Canada and the United States seem to disregard the concept of ecological fallacy in understanding the disparity for different groups of people, not just in the school effectiveness research. Some people read statistical tables containing data on the incomes of various categories of people and then interpret the evidence as applying to the individuals in those categories (Hacker, 1997).

For example, a recent news article claimed that “the rich are getting richer and the poor are getting poorer” (Thomas & Gross, 2002). Instead of being locked into the bottom category and receiving less and less income over time, substantial evidence, in fact, shows that the majority of people in the lowest income categories move into higher brackets as they gain education and experience, and some of the people in the higher brackets decline in income as they get older and retire (Sowell, 2009). While disparities in income are evident, some authors commit an ecological fallacy when arguing that income data gathered from statistical categories can be directly attributed to the individuals in those categories without acknowledging that the data are gathered at one level and interpreted at another level.

Recently developed multilevel methods are directed at properly assessing the cross-level interactions between individual variables and group variables by analyzing the effects at their proper levels. As noted earlier, students are influenced by classrooms managed by teachers within schools, districts, and provinces. As such, schooling is conceptualized as a hierarchical system with students at one level, teachers and classrooms at a higher level, schools at yet a higher level, and districts and provinces at levels above schools. Of course, when estimating the cross-level effects, the individual and school variables must be defined and measured at their corresponding levels. At present, there are a number of multilevel modeling procedures for assessing hierarchical models, and multilevel regression models, especially Hierarchical Linear Models, are most commonly used (Hox, 2002). In this dissertation, Hierarchical Linear Modeling (HLM) is used to examine the cross-level school effectiveness using the nested TIMSS 2007 data from British Columbia and Ontario.

An obvious advantage of multilevel modeling is that researchers can specify and test cross-level models because the multilevel modeling takes the hierarchical structure of the nested data into account, and each level is formally represented by its own sub-model. The results in this analysis are more interpretable than both disaggregate and aggregate analyses because the models allow researchers to calculate the proportion of the variances in the dependent variables that are attributable to each distinct level. Thus, school effectiveness is decomposed into individual student variance, teacher and classroom variance, school variance, district variance, and provincial variance. The most significant advantage of multilevel modeling is the increased accuracy of the estimated coefficients. Clustered data, by its nature, violates the homogeneity and independence assumptions that are basic in traditional linear analyses. In multilevel modeling, these violations are not a problem because separate equations are written for each group, and the homogeneity of variance and independence assumptions are most likely true within each group, but are not necessarily true between groups. Furthermore, a unique random effect for each group is incorporated into the statistical procedures, and the variability in the random effects is taken into account when the standard errors are calculated, which means that the effect parameters and the error variances are estimated separately, and for this reason, they are more realistic (Ethington, 1997).

As noted, Hierarchical Linear Modeling is much more appropriate than disaggregate or aggregate analyses for nested data. Multilevel modeling is also subject to a number of conditions that must be carefully examined. First, the accuracy of multilevel analysis depends on a few assumptions: 1) the dependent variable is assumed to be normally distributed; 2) collinearity among the independent variables is assumed to be low; and 3) all theoretically defensible levels are included in the analysis. In the preliminary analyses presented in chapters 3 and 4, it will be shown that the variables in this dissertation meet the assumptions: normal distribution of data and low collinearity. Second, these procedures are still problematic if poor contextual measures are constructed from arbitrary administrative or political units that are not meaningful contextual variables (Greeland, 2001). Finally, a large sample is crucial for multilevel analysis. In order to control for potential confounding factors and provide meaningful effect parameters, there should be at least 30 groups and each group should have at least 30 individuals. In school effectiveness research, because of the interest in the group-level effects, there should be at least 50 schools (Hox, 2002). One important reason that the Canadian TIMSS 2007 data set was used in this dissertation is because it includes large nested samples of students and schools from British Columbia, Alberta, Quebec, and Ontario.

The TIMSS study was designed to inform public institutions about the effectiveness of their curricula and instructional methods, and to help administrators, principals, and teachers improve their teaching as well as to help students improve their learning of mathematics and science. In this program, the achievement of students in mathematics and sciences at the fourth and eighth grades is assessed every four years, which began in 1995. To obtain the student samples, the TIMSS researchers have used a two-stage sampling procedure, within which a random sample of schools are first selected with a probability proportional to the size of the schools in the province, and second, one or two intact fourth and eighth grade classes of students within each of the sampled schools are selected. In 2007, the fourth cycle of data collection, four Canadian provinces, British Columbia, Alberta, Quebec, and Ontario, participated in the TIMSS study. Alberta administered the assessment only at the fourth grade. In this dissertation, eighth grade students from British Columbia and Ontario, the two provinces that mainly offer English programs, were selected. Additionally, only the mathematics scores were used as representing the students' academic achievement. Mathematics and science are most likely to be taught and learned in school compared to other subjects like reading and social studies, which are often learned and practiced, at least to some degree, at home. In the Canadian TIMSS 2007 data, the scores of mathematics and science are highly correlated, ($r=.97$, $p<.001$). Consequently, using the achievement score in mathematics or in science will have similar results.

In this research, besides the students' tests on mathematics and sciences, the students, their mathematics teachers and school principals were asked to complete questionnaires to obtain information on the home and school conditions the students likely experienced. In brief, students answered questions about their backgrounds, attitudes, and their home and school environments; the teachers responded to questions about their backgrounds, their attitudes, their teaching activities, the resources, and social environment in their classrooms. The principals, in turn, responded to questions about their school's resources, social climates, and school administration. A very important advantage of the TIMSS is that it obtained information from the teachers who actually taught the students in mathematics so the information about the students' learning environments is more accurate than in a number of other international studies. The Programme of International Student Assessment (PISA), for example, relies on the school principals' assessment of the teacher and classroom variables. Nevertheless, the TIMSS data set still has some limitations, which must be understood.

Limitations

There are, in fact, several limitations in a study using the TIMSS data. First, it is secondary data analysis. Existing data sets, like TIMSS, provide large representative samples, standard items, and standard indices that cover many topics, time periods, and countries. Yet, the cost for researchers to use these data sets is low, obtaining the data, preparing them for analysis, and conducting the analysis. Often these surveys are designed and implemented better than many local surveys. The TIMSS 2007 database is a wealth of data for secondary analysis that provides valuable indicators of student achievement, other student variables, teacher variables including teaching methods, and items that measure the administration of the schools. Still, other investigators collected the TIMSS 2007 data without the specific research question that a secondary analyst may have. Thus, my dissertation is limited by the data collected and definitions used in the TIMSS 2007 data set because the items from students, teachers, and principals were constructed by other researchers. Consequently, validity is of concern because the survey items may not be precise measures of the concepts, or the variables may be poorly constructed. Additionally, not every variable that I am interested is available in the data set. Nevertheless, for the research problem in this study, the advantages of this data set and the secondary analysis clearly outweigh the disadvantages.

Second, this study is cross-sectional which assesses the relationship between the school variables and the students' academic achievement and infers the relationship to the school and student populations. The large sample of the TIMSS data allows for the multilevel modeling of variables and the valid inference of the findings. However, in a cross-sectional study, the independent and dependent variables are measured at the same time. Therefore, this study does not allow for definitive conclusions about the causes and the effects of the students' education because the temporal sequencing of the relationship between the independent variables and the dependent variable is not examined. It is obvious that some variables, such as the schools' administration and the teachers' gender, are possible causes of variation in the students' achievement in mathematics because they have been decided before the students enter the schools and classrooms. Nevertheless, it is difficult to tell the time sequencing between the dependent and some of the other independent variables, for example, the homogeneity of students in the classrooms, the students' educational expectations, and their effort in completing homework. Additionally, this cross-sectional study does not provide entering achievement scores for the students, so it is difficult to evaluate how much change in the dependent variable has been caused by the independent variables. However, this dissertation assumed that virtually all the variation in mathematics scores can be attributed to the students, the teachers and classrooms, and the schools. Also, this dissertation assumed the stability of the covariate distributions over time, which is suspect, of course, because over time students, particularly adolescent students, change their attitudes and their behavior, sometimes dramatically.

The third limitation of this dissertation is that it is only possible to use two-level models with the TIMSS data, the school level and the student level. Ideally, a multilevel model of schooling would include the student level, the teacher and classroom level, the school level, the district level, and the provincial level. However, in the TIMSS data, the student samples were selected using a two-stage procedure, the schools were selected and then the students within those schools were selected. Consequently, there is only a two-level nesting of the data so that students are nested within schools. Even though the students could have been nested by their teachers and classrooms, there was only one, or, at most, two mathematics teachers and classrooms in each selected school, which results in a sample that is too small to construct a nested teacher and classroom level below the school level. Instead, the teacher and classroom variables were used as school-level variables. Consequently, the variance components attributable to the school and classroom levels are impossible to distinguish. Also, there is no school district level information in the TIMSS data, and thus the differences between districts cannot be evaluated. Nevertheless, this dissertation can compare the provincial differences by analyzing the samples of schools and students within each of the two provinces, British Columbia and Ontario. Still, missing a teacher and classroom level is an important limitation of this dissertation that needs to be recognized.

The forth limitation of this dissertation is that the restricted samples may reduce the generalizability of the findings. Specifically, this dissertation only analyzed the student and school samples within two provinces. Naturally, the findings of this dissertation can only be applied to other Canadian provinces with caution since there are significant differences between provinces in their educational policies, organization of school systems, and their students' achievement. Obviously, the results of this dissertation may not be applied to other countries where the educational systems are dramatically different from those in Canada.

A fifth limitation results because there is a large amount of missing values in the Canadian TIMSS 2007 data. In the school data set, 16 schools with 396 students in British Columbia and seven schools with 113 students in Ontario were missing. These missing schools were omitted from the analyses in this dissertation. The assessment of the missing schools found that students in these schools were significantly different from students in the other schools. For example, in both provinces, the students coming from schools with missing data had significantly lower achievement scores in mathematics than students from the other schools. Because there is no school level information, it is impossible to determine what might have caused the comparably lower student achievement in these schools. In the student data, the students' parental education had a relatively large amount of missing values, as high as 39%, which was probably caused by a number of students reporting that they did not know their parents' educational levels. The preliminary investigation found that in both British Columbia and Ontario, students who had missing information about their mothers' education were also likely to have missing information on other important background variables, such as their gender. Dropping these students has dramatically decreased the missing values of the student variables to below 10%. Unfortunately, it was found that in both provinces, the students who had missing values on their mothers' education, and were thus dropped from the analyses, had significantly lower mathematics achievement scores than the students included in the analyses. Not surprisingly, these students also had significantly less positive attitudes towards learning and lower expected educational attainment. Thus, using only part of the samples in both British Columbia and Ontario is a final limitation of this dissertation. Obviously, these limitations may affect conclusions in this dissertation. The limitation caused by missing values was examined carefully and the result is reported in chapters 3 and 4, and discussed again in Chapter 5.

Overview

In this chapter, I introduce the problem I chose to examine in this study. Simply, I examined the effect of school variables on the students' academic achievement when a number of student variables were controlled. More specifically, using the large Canadian TIMSS 2007 data set, the effect of four school variables and eight teacher and classroom variables that have been generally assumed to have impacts on the students' achievement were assessed while five important student background variables were controlled. Contradictory findings in the current literature on school effectiveness may be caused by inappropriate analytical procedures. With improved modeling procedures, this dissertation helps clarify some of the inconsistencies in the published research literature within the constraints of five limitations in this study.

In Chapter 2, I explain why and how researchers can study cross-level effect of schooling using multilevel modeling techniques. The first section explains the significance of using multilevel techniques in school effectiveness research after reviewing the multilevel nature of schooling and the methodological problems evident in the current research. I show that multilevel analysis is more appropriate than other methods for analyzing these educational data. The second section presents the multilevel modeling method, its advantages, and when its use is appropriate. Multilevel analysis, specifically HLM, is appropriate for the data analysis in this dissertation.

Chapter 3 is the methodology chapter. First, I provide a detailed description of the 134 schools, the 182 mathematics teachers, and the 2,392 students in British Columbia and the 169 schools, the 205 mathematics teachers, and the 2,080 students in Ontario that were included in this study. These samples were obtained from the Canadian TIMSS 2007 data set. Second, I present the dependent and independent variables and their theoretical significance, the items used to measure the variables, their descriptive statistics, and the chi-square and t-test results used to assess the significance of differences in these variables between British Columbia and Ontario. Finally, based on a review of literature and the descriptive statistics presented, I propose four hypotheses to be tested in this study.

In Chapter 4, I first review the analytical methods that were used to examine the missing values in some variables and collinearity among independent variables. Then, I report the procedures to test the four hypotheses. More specifically, this chapter compares the results produced by the three analytical methods (disaggregate analysis, aggregate analysis, and HLM), which demonstrates that the aggregate analysis gives the most biased results about the effectiveness of schools, including biased results for the effects of the schools' physical resources. In comparison, HLM gives the most accurate estimates using the nested TIMSS data set. Finally, I report the results of the HLM analysis. I found that the students' individual characteristics had the most important effects on their achievement in mathematics. The school and the teacher/classroom variables had relatively smaller effects. However, there were moderately large and significant effects from some of the institutional variables and they varied by province. The physical resources of the schools evaluated by the school principals and the classroom teachers had no positive impact on their students' academic achievement in either British Columbia or Ontario. Many of the findings in this study are consistent with the best research literature.

A discussion and interpretation of the results are presented in Chapter 5. In this chapter, I first discuss the limitations of this study caused by the missing values. Given the constraint of excluding some schools and students, the findings of this study could still be generalized to other provinces. I then discuss the findings in relation to the research literature, namely, the biases in an aggregate study, the student, teacher/classroom, and school effects, measurements of the physical resources, and provincial differences. I suggest that, currently, the contradictory findings about school effect may be caused largely by the analytical procedures that have been used by researchers. Finally, I suggest further research that could be conducted.

CHAPTER 2

ESTIMATING THE SCHOOLS' CROSS-LEVEL EFFECTIVENESS

This chapter explains why and how researchers can and should study cross-level effect of schooling using multilevel modeling techniques. In this respect, the chapter has two sections. The first section explains the significance of using multilevel techniques in school effectiveness research after discussing the multilevel nature of schooling and the methodological problem in the current research literature using nested educational data, mainly the aggregate analysis and an ecological fallacy caused by aggregation. For this reason, multilevel analysis is suggested as a more appropriate model for analyzing nested educational data. The second section presents the multilevel modeling procedure I used, Hierarchical Linear Modeling (HLM), its advantages, and when it is appropriate to use this method.

Cross-level Effectiveness of Schooling

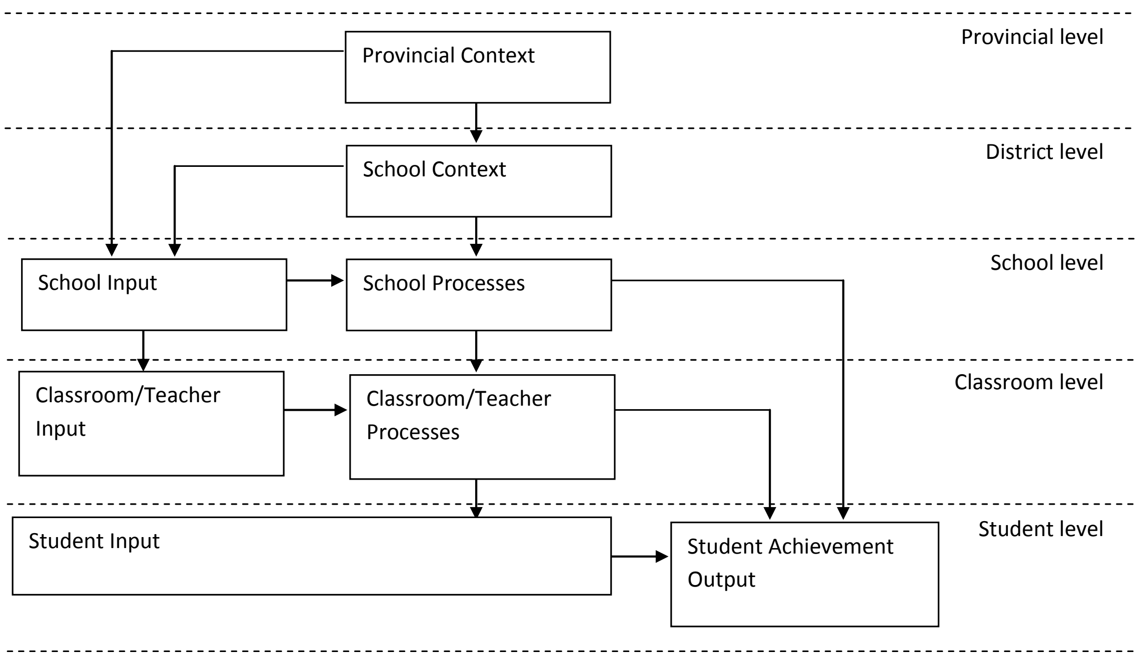
Cross-level effects exist in education because much of what goes on in schools occurs within groups and within hierarchical organizations; students are nested in classrooms located within schools, which are located within school districts, and so on. Researchers have long realized that the hierarchy in many educational data must not be ignored. In general, once the groupings are established, to various degrees they influence their members, and in turn, the groups are influenced by their group members, which implies that individuals from the same social context may behave more similarly than individuals randomly sampled from the population. In other words, observations based on these individuals are not fully independent. Similarly, the educational influence on students occurs in the groups to which the individual students belong. Although social scientists are aware of this cross-level effectiveness (Bronfenbrenner, 1976; Burstein, 1980; Cronbach, 1976; Mischel, 1977; Webb, 1977), which refers to the effects of higher-level organizational units (e.g., teachers, classrooms, and schools) on lower-level group members (e.g., students), it is still unclear how to best estimate the effects of the group contexts. Therefore, extensive theoretical and statistical work has been conducted to explain the complex influences of individuals and their groups on the individual's behaviour.

As a typical cross-level issue, school effectiveness—the influence of schools on the students' educational outcome—has been continually investigated even when the statistical methods were inadequate. In this literature, school effectiveness research (SER) asks the question about how much do schools differ in their students' performances, or more precisely, why are some schools more effective in helping their students learn than other schools, if the differences are not due to the students' own characteristics, for example, SES, motivation, expectation, etc., which vary within schools (Reynolds, Teddlie, Creemers, Scheerens, & Townsend, 2000). Obviously, in SER the effectiveness of schools is thought to be a cause of differential student achievement, within which some school characteristics lead to relatively higher performances when the characteristics of the students are held constant. Moreover, school effectiveness is multilevel because of the nested nature of schooling, within which students are grouped within classrooms or by teachers; classrooms or teachers are grouped within schools; and schools are grouped in school districts. It is noticeable that more than what is happening within schools and classrooms, the higher administrative levels (districts and provinces) also have some authority over some of the conditions that affect students. Briefly, each higher level in the system provides, to some extent, contextual conditions for the lower levels. Through direct and indirect cross-level effects, these hierarchical levels in the system eventually affect the students' learning.

This multilevel view of school effectiveness actually implies that school characteristics cannot be conceived as treatments uniformly administered to all students within a school, nor can the "treatment effect" be conceived as a constant increment to the development of each student within a specific school. As noted, these two implications are consistent with how schools operate and how students learn. First, in schooling, actors at each level of the organization decide how to distribute scarce resources under their control. Such resources include people, time, and materials (Bidwell & Kasarda, 1980). Thus, decisions at any level of the organization constrains the opportunities for actions at the lower levels. This means, for instance, that although two teachers may go to work each day in the same high school, the resources available to them for instruction may differ because one teacher is teaching a subject that is valued more by school administrators than the other. At the same time, the satisfaction they derive from their work may vary as a function of the students and the subjects they teach. The ultimate effect of this multilevel process is that the resources for instruction and, therefore, the opportunities for learning often vary across schools, classrooms, and eventually, students in very complex ways including the various levels which have control over how to distribute and use the resources. And second, given the same opportunities for learning, different students may respond quite differently. Naturally, an adequate conceptualization of the effects of schooling should model not only how schools differentially allocate resources that have an impact on the opportunities for students to learn, but also model how students differentially respond to the available opportunities because of their unique characteristics. These levels are illustrated in Figure 1.

Figure 1

The Cross-level Effectiveness of Schooling



There is a clear analytic consequence of this conceptualization of how schools work, but the procedures to estimate these effects are not obvious partly because though school effectiveness is primarily seen as an issue for individual schools, all relevant factors that are associated with the higher performance of students should be taken into consideration including cross-level effects and the interaction of these effects, specifically, the student-level variables, the classroom-level variables, the school-level variables, the district -level, and the provincial-level variables. Only by including key variables from each of the appropriate levels, are the processes and contextual factors in the model likely to be correctly estimated. Figure 1 illustrates the cross-level effectiveness of schooling and synthesizes the important variables of the schooling system in Canada and many other countries.

This cross-level model illustrates the design of school effectiveness research, which is to detect the impact of the relevant input variables on the output variables and to decompose the schooling process by revealing the variables measured at various levels that are having major effects on the students' achievement. In other words, the major task of SER is to find the association between the hypothetical effectiveness-enhancing conditions and the measures of output, which is usually calculated in terms of the students' achievement. More specifically, within this model, the *output* variables usually include the students' cognitive outputs (standardized tests in mathematics, reading, science, GPAs, etc.) and less often their non-cognitive outputs (student attitudes including their self-concepts, academic aspiration, study habits, dropout rates, graduation rates, jobs, or earnings, etc.). At the student level, the students' background, their motivation, and their aptitudes strongly determine their achievement. At the classroom/teacher level, the time and opportunity the students have to learn the subject, the quality of the instruction, the curriculum, the grouping procedures, and teacher behaviors have a significant impact on the students' learning. At the school level, most school factors, such as an orderly climate, adequate school resources, and the constant evaluation of student achievement, are variables that affect the quality of instruction, the time, and eventually, the opportunity students have to learn a subject. Thus, all these variables affect the students' learning. In brief, within this cross-level framework of school effectiveness, students, teachers (classrooms), and schools, as well as the contexts above schools are important levels that must be considered in understanding the variables that affect the students' learning. Generally, among variables within this multi-layered organization, the students' own characteristics are thought to be the most important.

Student Effects

It is common that within the same school, the same classroom, and with the same teachers, some students perform better than others. Furthermore, compared to the school and classroom/teacher impacts, the most influential impact on the students' academic achievement comes from their own backgrounds. Depending on the research methods, the percentage of variance in the students' achievement that is explained by the differences between students varies between 70% and 95% (Bosker & Witziers, 1996; Bryk & Raudenbush, 1992; Coleman, Campbell, Hobson, McPartland, Mood, Weinfeld, & York, 1966; Creemers, 1994; Rowe & Hill, 1998; Luyten, 1994; Madaus, Kellaghan, Rakow, & King, 1979; Scheerens & Bosker, 1997; Stringfield & Teddlie, 1989; Townsend, 2007). Consequently, only between 5% and 30% of the variance in the students' academic achievement can be explained by differences at the higher levels (schools, districts, or provinces, for example). As a result, the students' backgrounds should be included and analyzed in all SER research.

There has been a long tradition of psychological and sociological research showing that various individual-level variables affect the students' educational outcomes. Concerning the students' backgrounds, one well-examined theory is the Wisconsin model of educational and occupational attainment (see, for example, Alexander, Eckland, &

Griffin, 1975; Sewell, Haller, & Portes, 1969; Sewell & Hauser, 1975), which posits that the educational aptitudes, expectations, and aspirations of students affect their education and career outcomes. In this theory, social resources available to students are influenced by parental education, parental occupation, and family income, all of which have key influences on the students' educational and future career success. Additionally, social cognitive theory suggests that psychological dispositions are neither determined autonomously by individuals nor dictated entirely by environmental forces (Bandura, 1986); instead, these important dispositions are developed within an interactive system of individual actions, cognitions, other affective conditions, environments, and the students' experience. As such, the identification of individuals' variables is fundamental in most SER research literature. In addition, the teacher/classroom-level and school-level impacts are thought to be mediated by the student-level variables.

In general, when considering the cross-level effectiveness of schools, the *student input* variables are often the students' SES (e.g., family size, family income, parental occupation, parental education, family possessions, family educational environment, etc.), background (age, gender, grade, kindergarten attendance, aptitude, etc.), and their affective disposition (self-control, self-concept, self-esteem, motivation, educational expectation, etc.). Only after controlling for these individual variables, can the assessment of the impact of the teacher, classrooms, and schools be considered reasonably accurate.

Teacher, Classroom, and School Effects

In total, schools account for between 5% and 30% of the variance of the within-school variation in the achievement of students (Bosker & Witziers, 1996; Bryk & Raudenbush, 1992; Coleman, et al., 1966; Creemers, 1994; Rowe & Hill, 1998; Luyten, 1994; Madaus, Kellaghan, Rakow, & King, 1979; Scheerens & Bosker, 1997; Stringfield & Teddlie, 1989; Townsend, 2007). Among all schooling related variables, the teacher/classroom variables have the largest effects on the students' learning when the other background variables are controlled (Goldhaber, 2002; Hattie, 2009). Nye, Konstantopoulos, and Hedges (2004) reviewed seven studies and reported that the proportion of the variance in the students' achievement gains that result from the teachers' variables range from 7% to 21%. In multilevel studies, teachers account for more than twice the variation in the students' achievement changes than the school variables (Meyer, 2001). No wonder that after effectiveness of students' background variables, teacher effectiveness is reported to be the next largest set of variables affecting the academic achievement of students in schools (Rivers & Sanders, 2000). Moreover, the impact of teachers may vary both across grades and within grades for teachers who are teaching the same subject. As a consequence, a school's effect could actually represent an aggregation of the effects of the teachers in the various classrooms (Hanushek, Kain, & Rivkin, 1998; Meyer, 2001; Wayne & Youngs, 2003; Webster, Mendro, Orsak, & Weerasinghe, 1996).

Because teachers and classrooms have significant roles in the way students are educated, existing research has examined the aspects of the teacher/classroom quality that matter. Briefly, the *teacher input* variables often include teachers' age, gender, ethnicity, experience, educational level, major, teaching skills, and salary. At the same time, the *classroom input* variables often include the grade level, the heterogeneity or homogeneity of the students, class size, and the resources that are available in classrooms. Above the teacher and classroom inputs, the s*chool input* variables focus on a school's human and physical resources that are directly related to the quality of instruction, and eventually the students' opportunity to learn. As such, the most common school input variables are schools' size, expenditures, physical facilities (labs, lights, and building quality), instructional materials, the student/teacher ratio, administration, and the social norm.

Other than the inputs of classroom and school variables, cross-level school effectiveness research also must consider the processes in the learning environment that could potentially affect the students' learning. Namely, *classroom/teacher process* variables include the time the teacher spends on task, the structure of teaching, the opportunity that the teacher provides for students to learn the content compared to what is tested, the expectations that the teachers have for the students, the evaluations of the students' progress, and the reinforcement the teachers provide. Similarly, s*chool process* variables include those that would help build a more effective learning environment in classrooms, for example, the achievement-oriented policies of schools, school leadership, consensus among teachers, cooperation among teachers, quality of the school curricula, the authoritative and disciplined climate, and evaluation of the students' progress. Some of the school processes are, of course, partially or completely established by district and provincial regulations.

District and Provincial Effects

Above individual students and schools, there are noticeable differences in students' achievement between school districts and between provinces. Generally, Canadian students perform relatively well in the international surveys in mathematics, sciences, and reading, for example, in the PISA studies (Brochu, Gluszynski, Knighton, 2010). However, substantial provincial differences exist in the students' academic achievement. It is commonly recognized that Alberta, Quebec, and British Columbia generally have the highest academic achievement in Canada, with Ontario and Manitoba in the middle of the pack, while the Atlantic provinces and Saskatchewan tend to perform lower.

Table 1

Average Achievement Scores in PISA for British Columbia and Ontario

	Mathematics		Science		Reading	
Year	BC	ON	BC	ON	BC	ON
2009	524	527	535	531	525	531
2006	523	526	539	537	528	534
2003	538	530	527	515	535	530
2000	534	524	533	522	538	533

Note. Source of data: Human Resources and Skills Development Canada, Council of Ministers of Education, Canada, and Statistics Canada, 2010.

Table 1 compares the differences between British Columbia and Ontario, the two provinces included in this study, for their students' achievement on PISA tests in 2000, 2003, 2006, and 2009. It is observed that, in 2000 and 2003, students in British Columbia outperformed students in Ontario in mathematics, science, and reading, but in 2006 and

2009, students in Ontario outperformed students in British Columbia in mathematics and reading, and in science they performed almost equally well as students in British Columbia. These results clearly suggest that, as a context, the province may have an impact on students' performances. Moreover, evidence suggests that the impact is not static but constantly changing.

The provincial differences may result from differences in educational policies or expenditures. For this reason, Table 2 presents the yearly expenditure in elementary and secondary schools in British Columbia and Ontario for the school years from 2005–06 to 2009–10. It is noted that Ontario consistently spent more money on students than did British Columbia. It is reasonable to make a tentative connection between the student performances in Ontario and the increased expenditures in this province of 11%–16% more. Of course, it is also possible that the achievement may not be related to the expenditures.

Table 2

School Expenditures in British Columbia and Ontario

	Total expenditures		Full time enrollment		Expenditure /student		Difference
Year	BC	ON	BC	ON	BC	ON	
09–10	$4,878,898,951	$19,784,688,747	541,083	1,891,568	$9,016	$10,459	16.0%
08–09	$4,798,247,734	$19,192,563,461	542,356	1,910,585	$8,847	$10,045	13.5%
07–08	$4,678,963,136	$18,353,459,226	547,840	1,930,934	$8,540	$9,504	11.3%
05–06	$4,265,360,667	$16,911,848,973	567,523	1,944,030	$7,515	$8,699	15.8%

Note. Source of data: Ministry of Education, British Columbia and Ontario Ministry of Education.

While the cross-level effects of schooling are not difficult to understand, modeling these effects is difficult. Limited by knowledge, data, and other conditions, many studies have attempted to evaluate cross-level effects by using classical parametric procedures. Logically, there are two procedures available to evaluate multilevel effects at one level: either the disaggregation or aggregation of the variables. The former distributes the higher level variables to the lower level and thus the higher order variables become, for the purposes of analysis, individual attributes. The lower level is then used as the unit of analysis. The latter aggregates the data to the higher level and uses the higher level as the unit of analysis. Unfortunately, both of these procedures contain flaws in the specification and estimation of the effects.

In the disaggregation procedure, the analysis of the variable distribution is not appropriate because it ignores the fact that the data violate the basic assumptions of parametric analyses: the normal distribution of the data, the homogeneity of variance, and the independence of the observations, specifically. Ignoring the nested structure in the data would produce standard errors that are misleadingly small, and consequently the confidence intervals would be deceptively tight (Paterson, 1991). More importantly, this method produces inflated Type I errors in testing the effects of the independent variables, which implies that the null hypothesis is rejected more often than expected. Nevertheless, even though the disaggregation makes no distinctions between the between- and the within-group variances, it may produce meaningful coefficient estimates. A more significant problem exists when the data are aggregated. Actually, aggregate analysis is more commonly used in school effectiveness research because they are simpler, and often the data are readily available. But, there are serious problems with these analyses as I show in the next section.

Aggregate Modeling and the Ecological Fallacy

In aggregate analysis, often referred as ecological analysis, data on individuals are aggregated to higher levels, such as classrooms or schools. Then, the relationships among the aggregated variables are investigated. There are two major types of measurement of aggregates: summaries of the distributions of individuals within the aggregates and ecological variables that are defined directly at the aggregate level. Most dependent variables are summaries of individual-level variables, such as the students' achievement scores or their attitudes, and the ecological-level analysis is expressed as the average individual outcome on the ecological unit, such as school-level achievement means. Contextual variables may be direct measures at the aggregate level, such as the schools' physical resources or funding, or summaries from individuals, such as the schools' average socioeconomic status (SES) of students.

There are an impressive number of quantitative ecological studies that relied on ecological correlations to make inferences about the behavior of individuals. And, in school effectiveness studies, the ecological correlation continues to play an important role in identifying the relationship between two variables for groups of students. In a recent educational report from the Atlantic Institute for Market Studies (Laurie, 2007), for example, the scores that students received on standardized tests and the grades they received from their teachers were both aggregated to the school level, and these variables were then correlated. As a result, the study concluded that there was a link between grade inflation and the students' performance on the standardized tests (Laurie, 2007). Another commonly seen example is the relationship between a school's average achievement score and the average SES of students (Marks, Cox, & Pomian-Srzednicki, 1983, 1985). In these aggregate or ecologic analyses, researchers often assume that the relationship at the group level is the same as it is at the individual level. Unfortunately, this assumption is not always true because of the disparity between the ecological- and the individual-level relationships. Equating relationships at these two levels will, unfortunately, result in an ecological fallacy.

An Ecological Fallacy

Conceptually, an ecological fallacy refers to the incorrect assumption that the relationships between variables observed at an aggregated, or ecological, level are necessarily the same as the relationships between the same variables at an individual level. Depending on the causal structure assumed by the researchers, the aggregation of data and then the attribution of the results to individuals can either confound the effects among variables or produce spurious relationships. In this respect, ecological fallacy is also called ecological bias, cross-level bias, or aggregation bias (Greenland, 2001; Greenland & Morgenstern, 1989; Greenland & Robins, 1994; Morgenstern & Thomas, 1993), which all refer to the difference between the coefficients obtained at the aggregate level and the coefficients of interest that would be attained at the individual level if they were actually calculated.

According to Robinson (1950), who had a most dramatic impact on our understanding of an ecological fallacy (a term coined by Selvin, 1958), the term ecological fallacy points out that the pattern of correlations displayed at the ecological level cannot be simply generalized to the behavior of individuals. Specifically, ecological coefficients are usually much higher than individual coefficients. Furthermore, as the number of aggregate units decreases, the ecological coefficient increases in size, and as this happens it becomes clearer that the analysis from the aggregate level cannot be applied to the individuals who compose the aggregates. From a conceptual perspective, as Oakes (2009) indicates, group-level phenomena are never simple aggregations but are very complex, dynamic, and multilevel phenomena. As such, group level, or ecological, analysis should not be equated with individual-level analysis.

The statistical perspective also prohibits the simple assumption that the group level and the individual level relationships are identical. First, the aggregated measures and the subsequent analysis have questionable measurement errors. Most often, as summaries of the individual distributions, the means are used to represent the group-level measures at the ecological level, which are extremely vulnerable to outliers. Even without extreme outliers, the means cannot be seen as unbiased measures of the group members because there are numerous possible distributions of individual scores that can all share the same aggregate group mean. Consequently, analysis based on the aggregated measurements can hardly reveal the true relationship at the individual level. Second, even with normally distributed individual measures, the group-level relationships still are not equivalent to individuals because of the numerous possible relationships at the individual level. Only when the relationships between the variables are identical in all groups, can the individual relationships be inferred accurately from the aggregate relationships. Unfortunately, this condition is very rare.

Consequently, there are many possible differences between the individual-level and the group-level analyses in which the ecological data can give misleading results for individuals. If a researcher assumes that all other possibilities can be ruled out except for the ideal one that both the individual and the ecological relationships are identical, the researcher could reasonably conclude that the ecological and the individual-level analyses are similar. For example, Gove and Hughes (1980) manipulated their data to meet this criterion. But, as Handel (1981) pointed out, there are still many possible relationships, and it is difficult to examine all these possibilities. For this reason, the information lost during the aggregation of the data makes it almost impossible to detect the relationships between variables among individuals. Essentially, aggregation causes significant difficulties in analysis, which becomes the origin of an ecological fallacy.

Causes of an Ecological Fallacy

Often, both ecological- and individual-level analyses have the same focus for their inferences, the effects at the individual level. As we already know, in the individual analysis, these target effects are at the same level as the unit of analysis while at the ecological level, these target effects are at a lower level. Unlike an analysis at the individual level, an ecological analysis at a level above the individual level (e.g., school, district, and province) cannot link the individual outcome to an individual's exposure to the context, nor can it link one individual to another individual. A number of reasons are responsible for this broken linkage and the consequent aggregation bias. Among these reasons, the difficulty to identify and control for confounding factors is the most fundamental, but there are other reasons such as the change in the meaning of variables from one level to the other as well as the measurement error.

Confounding. The presence of summary measures in an ecological analysis introduces a major source of uncertainty to its inference. The effects of summaries depend on the joint individual-level distributions within the aggregates, but the summaries often do not fully determine, and sometimes do not even seriously constrain, the joint individual-level distributions largely due to confounding variables. One complexity of having confounding variables is that they can happen both at the individual and at the contextual level.

At the individual level, ecological estimates will be biased if individuals are not affected equally by confounding variables across groups. It is possible that there are positive confounding effects in certain groups which cancel negative confounding effects in others. Even if the within-group effect is equally confounded by one unmeasured variable in every group, the ecological estimates can still be biased. One of the reasons is that the ecological data usually refer to arbitrary administrative groups, such as schools or classrooms, which may not accurately represent the social context of each group member, such as the student's SES, abilities, or attitudes. In other words, the confounding variables are not similarly distributed among all group members. This gives considerable opportunities for confounding variables to bias the results in ecological analysis (Greenland, 2001).

At the contextual level, the ecological bias can also occur when unmeasured contextual confounding variables are differentially distributed across groups, even if they are not confounders at the individual level (Greenland & Morgenstern, 1989). As commonly known, students are often exposed to both identifiable and non-identifiable school variables. The contextual confounding variables may change the students' outcomes in various directions and to various extents, thus contaminating the effects of the explanatory contextual variable under study. For example, in an aggregate study aimed at examining the relationship between the students' exposure to the school facilities, the students' mean score may be affected by uncontrolled contextual variables, such as the longer school-wide instructional time, or higher academic press from teachers. Thus, conclusions about the impact of one school context variable cannot be made because of these unidentified and uncontrolled confounding factors.

Besides, there are possible confounding variables that may exist at both the ecological and individual levels and they may not be causally connected to each other. Consequently, the condition of having no confounding variables at an ecological level cannot guarantee that there are no confounding variables at an individual level. The aggregation of data makes it impossible to identify and control for confounding variables, and to complicate this even more, the confounding effects can be caused by the fact that the contextual and individual variables are confounded in the aggregated data, which means that an irrelevant confounding variable at one level may be introduced at another level.

Moreover, because of confounding variables, it is inappropriate to use statistical operations to control for individual level confounding variables at the ecological level. Suppose that we have a confounding variable, X, and the individual-level data allows the control of this variable in a straightforward way by the stratification of X. In contrast, at the ecological level, due to the uncertainty of this confounding variable, its control cannot be accomplished by only using the aggregated data. For example, the students' SES cannot be adequately controlled by the group mean SES because it may not properly represent all the students in that group.

In general, ecological data alone tell us little about either contextual or individual effects in the social sciences, including education, mainly because of the possibility of confounding variables and the lack of data at the individual-level to detect them. Lacking within-group data on joint distributions due to aggregation, an ecological analysis must necessarily rely on prior information for inferences about the contextual effect. Unfortunately, such prior information may not always be available, or contradictory interpretations of the effects may exist. In fact, this is a key controversy in ecological studies. Thus, ecological studies that adjust for the confounding factors must either employ external data about non-identified individual variables, or invoke questionable assumptions about the relationships among variables. The unidentified nature of confounding variables means that neither approach can be fully tested in ecological data. Nevertheless, this identification problem, which is an absolute demarcation between ecological, or aggregate, and individual studies, continues to be misunderstood or ignored by many social scientists (Greenland, 2001).

The unidentified nature of confounding variables is so fundamental to an ecological fallacy that many other reasons for the fallacy can be related to confounding variables. For example, a special confounding effect, grouping, happens when no clear grouping principles are used; in other words, the confounding is not identifiable. For example, Hannan and Burstein (1974) used two variables to show that grouping randomly results in unbiased estimators, but with very low efficiency, while grouping by values of the independent variable yields unbiased estimators with relatively high efficiency. Conversely, grouping by values of the dependent variable results in biased estimates. If grouping by a third variable that is not explicitly included in the substantive model, as often happens, there is no bias in the results if either 1) a researcher groups by a variable causally unrelated to X (net of Y), or 2) a researcher groups by a variable that is causally unrelated to Y (net of X). However, if both conditions happen simultaneously, the results will be biased. It is noticeable that grouping may be used as a way to control for confounding variables, but we must remember that grouping ordinarily magnifies the effect of specification errors at the individual level. Furthermore, it is impossible to use grouping to control for all of the possible confounding factors. Besides the confounding variables, there are other reasons for an ecological fallacy, which are less fundamental, but are still so important that they cannot be ignored.

Change of meaning after aggregation. The meaning of an aggregate variable may be changed after the aggregation from individual level data. As such, an aggregate variable may measure a different construct than its namesake at the individual level. Moreover, the aggregate-level variable taps more constructs than its corresponding individual-level variable. For example, Cronbach (1976) noted that if an individual is college-educated, it indicates what he would be inclined to purchase or what jobs he would be capable of holding. However, an aggregation of college education in a community not only describes an aggregate market and an aggregate employee pool, it tells what goods and services probably are well-supplied in the community and the kinds of jobs that are available. When aggregated and individual scores measure different constructs, the bias of using aggregated measures to investigate the pattern among individuals is inevitable (Firebaugh, 1978).

Measurement error. Measurement error can also have profound consequences for ecological studies simply because most covariates are rather crudely measured. Besides, the effects of measurement error on ecological estimates are not predictable partially because they can interact with confounding variables to produce serious bias in the results (Greenland, 2001; Greenland & Robins, 1994). In addition, the effects of measurement errors on the ecological- and the individual-level analyses can be quite different. In ecological analysis, it is often assumed that the ecological covariates are summaries of the individual measures. As a result, the ecological errors are assumed to be the same as the individual-level errors. However, the ecologic covariates are subject to errors beyond random or individual-level measurement errors. Some studies have examined the impact of such ecological measurement errors on cross-level inferences, but more research is needed, especially for situations in which the grouping variable is an arbitrary administrative unit other than a meaningful contextual variable (Wakefield & Salway, 2001).

Besides these problems that contribute to an ecological fallacy, we may need to consider intra-group confounding variables that may cause errors in the measurement (Greenland & Morgenstern, 1989), and we may need to consider selection biases (Morgenstern & Tomas, 1993). Because of the existence of these additional sources of biases, their possible measurement errors, and the loss of considerable information during the averaging process, especially when averaging over a multivariate distribution, when analyzing aggregate data, we not only lose the ability to extend inferences reliably to individual data, but we often lose the ability to estimate the direction and the magnitude of biases on the outcome measure. Greenland and Robins (1994) noted that biased evaluations can be especially difficult in ecological studies because of many potentially interacting covariates that may differ across groups. Furthermore, we cannot rely on the addition of more grouped data to eliminate the bias (Piantadosi, 1994).

Given these weaknesses of ecological analysis, it is almost impossible to infer individual effects from ecological analysis with much confidence. Nevertheless, some researchers have sought to modify the strict prohibition against downward cross-level inferences, and their most important conclusion is that aggregate data do not always yield biased estimates of individual-level unstandardized regression coefficients (Goodman, 1953; 1959). This line of research has, in fact, discovered some statistical conditions when there may be equivalence between the aggregate- and individual-level relationships, which are thought to be solutions to aggregate inferences (Subramanian, Jones, Kaddour & Krieger, 2009).

Three Tentative Solutions to Aggregate Inferences

In some educational and social science research, individual-level data are not available. Consequently, it is argued that the aggregate data and analysis are better than no analysis. Furthermore, encouraged by Goodman (1953, 1959), some statisticians have been searching for situations where valid downward cross-level inferences can be made with aggregate data. Up until the 1990s, three basic approaches had been attempted: homogeneous areas, the method of bounds, and ecological regression (Kousser, 2001).

King (1997) has combined the method of bounds with simulation procedures to produce a fourth technique, which has come to be known as EI—ecological inference procedures.

All these methods are based on making assumptions where realistic ecological inferences seem to logically apply to individuals. The homogeneous areas assumption is the simplest, which assumes that group members are grouped by the independent variable of interest. In other words, they are homogeneous. For example, if we need to find the relationship between the SES and the achievement of students, we select groups that are highly clustered by SES. In this way, the relation assessed at the group level can be inferred to the individual level. But, we cannot select such groups without committing a selection bias, which will influence the estimated effect. Additionally, it is often problematic to assume that every individual in a group will be affected in the same way by the independent variables because individuals differ on a number of important variables, such as their gender and psychological dispositions. Thus, this procedure results in unreliable and biased estimates.

The method of bounds calculates the logical bounds for each group and weights them. To overcome the problem that the logical bounds may be too wide to be of much substantive interest, Shively (1969) suggested that the bounds for some cells in contingency tables might be shrunk by making two assumptions: 1) about the direction and strength of the relationships among the internal cell entries, and 2), about the unknown probability distributions of the values within the range of estimates of these cells. Weak tests can be made by observing whether the calculations produce logically impossible effect parameters, which means whether the effect parameters seem reasonable from what we already know from the research literature. As indicated, these two assumptions can have very powerful influences on the estimates. To compensate for this, we can alter the assumptions and conduct different analyses to see how robust the estimates are to changes when different assumptions are made. Unfortunately, these estimates are often not very robust.

Goodman (1959) developed an ecological regression (ER) procedure, which addresses this problem but only for dichotomous variables. His method has different assumptions from the two previous procedures. Unlike the homogeneous areas method, ER does not assume that groups have the same effects, and unlike the method of bounds, it does not take the logical bounds explicitly into account. However, when using Ordinary Least Squares (OLS) procedures to estimate the parameters, the ER method has the crucial assumption that the error variance is random and the people in every group have the same distribution. This assumption is known as the constancy assumption of simple ER. If the groups and group members have different distributions, the assumption is violated and the estimates are biased.

King (1997) proposed a solution to this problem with the ecological inference (EI) procedure, which also only applies to dichotomous dependent variables. EI combines the method of bounds with Goodman’s regression technique, and estimates the effects using Maximum Likelihood procedures and numerical simulation assuming a bivariate normal distribution for the parameters. According to King, the basic model is robust to the aggregation bias, and produces realistic ecological estimates. Moreover, all the components of the proposed model are, in large part, verifiable by using diagnostic tests. In the basic model of EI, Outcome = individual factors + group factors + interactions of individual and group factors + error. This model suggests that including a random coefficient in the analysis may be better than using OLS procedures for aggregate data. In fact, King (1997) showed that EI represents a genuine advancement in making ecological inferences because it incorporates two elements, the method of bounds and varying parameters, which brings a new degree of accuracy to the analysis of aggregate data.

Nevertheless, similar to the previous aggregate methods, EI also has a fundamental and serious weakness because it is appropriate if—and only if—the three previously mentioned assumptions are correct: 1) the parameters are assumed to be distributed according to a truncated bivariate normal distribution; 2) the parameters are assumed to be uncorrelated with the regressors (in other words, aggregation bias is not present); and 3) the data do not exhibit any spatial autocorrelation (King, 1997). Since the basic model is not appropriate for every instance of aggregate data, it is necessary to determine how robust the model is when the data deviate from these three assumptions. However, researchers often have no idea whether or not these assumptions are correct. To a limited extent, one can evaluate the suitability of EI's assumptions for the data by using some special diagnostic procedures. However, the diagnostics do not always signal problems in the model even when they exist. Alternatively, the diagnostics sometimes point toward a poor model fit when the estimates are actually quite reasonable. With the limited utility of the diagnostics, EI does not bring us much closer to reasonably analyzing the underlying structure of aggregate data. Of course, we can justify the models we use by theory, but we need very strong theory to create the aggregate variables and build realistic analytical models. Still, without formal methods of determining how well the data fit the model, a researcher is left with a wide variety of reasonable estimates and no way of assessing whether or not any of them are appropriate. Obviously, the aggregate analysis may be useful merely because it may provide many model specifications that can be tested empirically.

As we see, both disaggregate and aggregate analyses try to model the cross-level data by arbitrarily flattening nested data into one level, which potentially brings considerable uncertainty into the estimates. These uncertainties essentially cause disparities between ecological and individual relationships, a problem that cannot be overcome by just using aggregate data. In school effectiveness research, a better solution is necessary so that nested data are analyzed in ways that avoid the errors caused by either the disaggregation or the aggregation of data. Multilevel modeling has emerged as such a method in which the data at their original levels are analyzed without either disaggregation or aggregation. The basic idea of multilevel analysis is relatively simple. Since the conceptual model of the schooling effects is multilevel (Barr, Dreeben, & Wiratchai, 1983), the statistical model must also be multilevel (see also Burstein, 1980; Goldstein, 1997; Raudenbush & Bryk, 1986; Rogosa, 1978). Consistent with the multilevel view of schooling, the multilevel models specify how the explanatory variables measured at each level of aggregation influence the distribution of outcomes at the student level (Burstein, Linn, & Capell, 1978). As a result, in multilevel analysis, the effects of schooling will not be manifest solely by variations among school means, but rather, they will also account for variations among the students attending the same school (Brown & Saks, 1981). The development of this conceptualization and measurement procedure has given multilevel modeling considerable advantages over both disaggregate and aggregate analyses. In the next section, I explain the conceptualization and procedures that are emerging for modeling nested data, which is followed by a discussion of the advantages of using these procedures.

Multilevel Modeling of School Effectiveness

Some recently developed multilevel methods are directed at properly assessing the interactions between individual variables and group variables. As noted in the school effectiveness research, students are believed to be influenced by classrooms managed by teachers, schools, districts, and provinces, to which they belong, and the properties of these groups are believed to be influenced by their students. As such, schooling is conceptualized as a hierarchical system with students at one level and schools at a higher level, districts at an even higher level, and provinces at the highest level. Naturally, the individual, school, district, and provincial variables are defined at their corresponding levels. Thus, in a two level analysis with students and schools, school effectiveness is decomposed into the school variance and the individual variance, a process that produces more accurate estimates of both the individual and the institutional effects. Equally important, the results are more interpretable than those produced by the previously discussed aggregate and disaggregate procedures.

At present, there are a number of multilevel modeling procedures that assess nested data, or hierarchical covariance structures (Hox, 2002). The multilevel regression model is the most commonly used and it is basically a set of regression equations at hierarchical levels. As such, it is also called Hierarchical Linear or Random-Coefficient Models. In fact, because Hierarchical Linear Modeling (HLM) is such a common procedure for multilevel modeling, in some writings, these two concepts are used interchangeably. Essentially, multilevel analysis can be used to examine various research problems by analyzing data with respondents nested within sampling units, such as data from factorial surveys, or longitudinal, and growth data, which makes them suitable to use in assessing school effectiveness. As a particular technique within the broad hierarchical analysis family, HLM, specifically, has a number of advantages over the more conventional disaggregate and aggregate modeling techniques. This can be explained by first examining the way HLM evaluates nested data.

The General Model and Sub-Models of HLM

The general model of HLM. At this point, a general two-level linear model of school effectiveness can illustrate the way HLM works. Basically, it is assumed that each of the levels has a different regression model represented by a simple regression equation, with its own intercept and its own slope. Because the students and the schools, at both levels, are sampled, we can assume that the intercepts and slopes are a random sample from a population of group intercepts and slopes, which is the definition of a random-coefficient regression model. In the simplest case, suppose that we want to study the impact of an individual student indicator X and a school indicator W on the students' outcome Y assuming that the students are nested within schools. The linear regression models can be written at both the student level, or level 1, and the school level, or level 2.

Level-1 model (within-group). The i_{th} student from the j_{th} school has an attainment:

$$Y_{ij}= \beta_{0j}+ \beta_{1j}X_{ij}+ r_{ij} \qquad \text{Equation 2.1}$$

β_{0j}—the j_{th} school's mean attainment, allowed to vary across schools

β_{1j}— the relationship between X and Y in the j_{th} school, which is allowed to vary across schools

X_{ij}—the X value of the i_{th} student from the j_{th} school centered at the grand mean of X

r_{ij}—a random error associated with the i_{th} student from the j_{th} school, which is assumed to be normally distributed with a mean of zero and variance of σ^2, or $r_{ij} \sim (0, \sigma^2)$

The parameters of equation 2.1 are then estimated for each school. At the same time, the variance of the parameters across schools is calculated and tested using a chi-square statistic to determine if they are different statistically or if the variability is due to chance. If there is non-random variability in the estimates, a level-2 model (between-schools) is estimated. In HLM, at level-2, we assume that the intercept and the slope of level-1 depend linearly on the contextual variable.

Level-2 model (between-group). Let W be a school indicator variable that is hypothesized to affect the school average attainment (β_{0j}) and the X-Y relationship (β_{1j}) within each school:

$\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + \mu_{0j}$ Equation 2.2

$\beta_{1j} = \gamma_{10} + \gamma_{11} W_j + \mu_{1j}$ Equation 2.3

W_j—W centered at the mean of all schools

γ_{00}—grand mean of Y

γ_{01}— difference in the school means of Y caused by W

γ_{10}—grand mean of the X-Y relationships

γ_{11}—difference in the school means of the X-Y relationship caused by W

μ_{0j}—random error in the j_{th} school's Y values from the grand mean of Y

μ_{1j}—random error in the j_{th} school's X-Y relationship from grand mean of the X-Y relationship

Now we can see from the general two-level equations presented above, if there are more levels, there are more nested linear models. Variables that are not measured are included in the error terms of these linear models. Using the variance component models, the disturbances can be separated into a school component and a student component. The student component is independent of the other students within one specific school, and the school component is independent of the other schools.

Single equation. Combining the level 1 and level 2 model yields a single prediction equation for the student achievement outcome:

$$Y_{ij} = (\gamma_{00}+ \gamma_{01}W_j+\mu_{0j}) + (\gamma_{10}+ \gamma_{11}W_j+\mu_{1j})\ X_{ij}+ r_{ij} \qquad \text{Equation 2.4}$$

After rearranging the symbols we obtain:

$$Y_{ij} = \gamma_{00}+ \gamma_{01}W_j + \gamma_{10}\ X_{ij} + \gamma_{11}W_j\ X_{ij} +\mu_{0j} +\mu_{1j}X_{ij}+ r_{ij} \qquad \text{Equation 2.5}$$

This equation shows why the standard regression analysis using either aggregate or disaggregate data is not appropriate for interpreting nested data. Typically, in a linear model using Ordinary Least Squares (OLS), the random errors are assumed to be independent, normally distributed, and have a constant variance. The random error in the outcome equation 2.5 has a more complex form, ($\mu_{0j} +\mu_{1j}X_{ij}+ r_{ij}$), where r_{ij} is a random variance among the individuals. But the errors $\mu_{0j} +\mu_{1j}X_{ij}$ depend on both μ_{0j} and μ_{1j}, which vary across schools and across the individuals' explanatory variable, X. It is also obvious that modeling the regression slopes using other variables results in adding interactions to the model ($\gamma_{11}W_j\ X_{ij}$). However, this coefficient conceals how the complicated error components are calculated by modeling the varying slopes across schools.

Reliability. The reliability of level-1 coefficients, analogous to those calculated by classical measurement theory, can be calculated by the mean reliability across all schools:

Reliability of $\beta_q = \frac{1}{j}\sum_{1}^{j} \tau qq \,/(\tau qq + vqqj)$ Equation 2.6

τ_{qq}—parameter variance

v_{qqj}—error variance

As with classical measurement theory, the reliabilities will be close to 1 when the parameters vary substantially across the level-2 units, or when the level-2 sample size (j) is large (Ethington, 1997).

This general model demonstrates all the fundamental statistical features of HLM. In empirical studies, depending on the research questions, researchers often use sub-models that are modifications of this general model. Commonly used sub-models include one-way ANOVA with random-intercept effects, one-way ANCOVA with random-intercept effects, and random coefficient models.

One-way ANOVA with random-intercept effects. If there are no explanatory variables at either level (β_{1j}=0, γ_{11}=0), the model will be a one-way analysis of variance (ANOVA) with random effects, which is a useful preliminary step in hierarchical data analysis. It is a random-effects model because the group effects are construed as being random:

Level-1 model: $Y_{ij}= \beta_{0j} + r_{ij}$ Equation 2.7

Level-2 model: $\beta_{0j} = \gamma_{00} + \mu_{0j}$ Equation 2.8

This model produces the school mean γ_{00} and its confidence interval, and it also provides estimates of the outcome variability at both levels. The variance of r_{ij} is the within-school variance component and the variance of μ_{0j} is the between-school variance component. From this one-way ANOVA, a useful parameter, the intra-class correlation (the amount of variance in the individual level responses that can be explained by the group level properties) or between-school variance, can be calculated:

$\rho = \text{Variance}(\mu_{0j})/\text{Variance}(\mu_{0j} + r_{ij})$ Equation 2.9

One-way ANCOVA with random-intercept effects. If there are only level-1 predictors and no level-2 predictors, and we constrain the slope X to be the same for all schools, the resulting model would be an analysis of covariance (ANCOVA) with random effects and level-1 predictors as covariates:

Level-1 model: $Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + r_{ij}$ Equation 2.10

Level-2 model: $\beta_{0j} = \gamma_{00} + \mu_{0j}$ Equation 2.11

$\beta_{1j} = \gamma_{10}$ Equation 2.12

The extension of the random-effects ANCOVA allows the introduction of level-2 covariates, and by adding a level-2 predicator, W, the model becomes:

Level-1 model: $Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + r_{ij}$ Equation 2.13

Level-2 model: $\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + \mu_{0j}$ Equation 2.14

$\beta_{1j} = \gamma_{10}$ Equation 2.15

In these sub-models, all the level 1 slopes (β_{1j}), or the effect of X on students, are constrained to be the same, and we can also construct a model that allows the slopes to vary from school to school, providing estimates of the school effects.

A Random-coefficients model. In a random-coefficients model, the level-1 slopes, or the X-Y relationships, are conceived to vary randomly across schools. In a simple random coefficients model, both the level-1 intercept and the level-1 slopes vary where there are no school predicators:

Level-1 model: $Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + r_{ij}$ Equation 2.16

Level-2 model: $\beta_{0j} = \gamma_{00} + \mu_{0j}$ Equation 2.17

$\beta_{1j} = \gamma_{10} + \mu_{1j}$ Equation 2.18

This three random-coefficient model allows us to estimate the variability in both the intercepts and the slopes across schools. Given one school predictor, W, we can use the full level-2 model of HLM in equation 2.2 and 2.3. In this model, both the intercepts and the slopes are predicted by school parameters. Conceptually, this model can be expanded to incorporate multiple Xs and Ws. At the same time, the school variables used to predict the intercept and each slope of Xs can be the same or different. For example, suppose we have two level-1 predictors, X_1 and X_2, and two level-2 predictors, W_1 and W_2, we can express the level-2 equations as:

Level-1 model: $Y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + \beta_{2j}X_{2ij} + r_{ij}$

Level-2 model: $\beta_{0j} = \gamma_{00} + \gamma_{01}W_{1j} + \gamma_{02}W_{2j} + \mu_{0j}$

$\beta_{1j} = \gamma_{10} + \gamma_{11}W_{1j} + \mu_{1j}$

$\beta_{2j} = \gamma_{20} + \gamma_{21}W_{2j} + \mu_{2j}$

In following this logic, there are numerous ways to model the data in testing a variety of research questions. For a more detailed explanation of these possible sub-models, see Raudenbush and Bryk (2002). From the argument presented above, it is clear that modeling nested data is too complicated for disaggregate or aggregate analysis to produce accurate estimates. In fact, advantages of multilevel modeling are emerging in the research literature because HLM is designed specifically to deal with the complexity of these types of nested data.

Advantages of Multilevel Modeling

As we can see, when modeling cross-level effects, multilevel modeling techniques, specifically HLM, evaluate the impact of the group contexts on the individuals while also accounting for individual differences. Compared to aggregate analysis, these techniques consider that the individuals within the same group are different and thus react differently to the group context. Compared to disaggregate analysis, these techniques consider that the individuals are more similar to others within their group than to individuals in other groups. This has given multilevel modeling a number of advantages over both disaggregate and aggregate analyses, and these advantages are especially important for school effectiveness research.

First, an obvious advantage of multilevel modeling is that researchers can formulate and test cross-level models without worrying about the unit of analysis when they have clustered data. Multilevel modeling takes the hierarchical structure of the data into account and each level is formally represented by its own sub-model. These sub-models express relationships among variables within a given level, and specify how variables at one level influence the relationships between variables at another level. As such, the data enter the model at their particular level, and the models allow researchers to investigate both lower level and higher level variance in the dependent variables. In other words, researchers can appropriate both individual and group variances in explaining individual outcomes, such as the academic achievement. Therefore, multilevel modeling overcomes the disadvantages of the disaggregation of group variables to individuals and the aggregation of individual data to groups.

Second, using multilevel analysis, researchers can estimate the variance and covariance components with even unbalanced nested data because the procedure enables the decomposition of individual variation into within- and between-group components. The within-group component, of course, describes the relationship inside groups, which allows for the comparison of individuals who are in the same groups, such as classrooms or schools. The between-group component takes account of differences between groups, which enables the comparison between classrooms or schools. Normally, the first step in multilevel analysis is to separate the within- and the between-group variance components. It is common to find in HLM analysis of school effectiveness that about 20% of the variance of the students' academic achievement is between schools, and about 80% of the variance is within schools. Multilevel analysis, in fact, identifies substantial differences between schools that conventional analyses would not detect because they do not partition the variance into within- and between-group components.

Finally, the most significant advantage of multilevel modeling is the increased accuracy of both regression coefficients and standard errors. Clustered data, by its nature, violates the homogeneity and independence assumptions, which are basic assumptions of OLS. However, these violations are not a problem in HLM. Even though HLM requires a few basic assumptions to be true about the data (normality, linearity, and equal interval, for example), which are similar to the assumptions required in conventional parametric analyses, the requirements for homogeneity and independence are adapted to suit multilevel data because HLM estimates separate equations for each group, and the homogeneity of variance assumption may be true within each group, but it may not be necessarily true between groups. Furthermore, a unique random effect for each group is incorporated into the statistical procedures, and the variability in these random effects is taken into account when estimating the standard errors, which means that the effect parameters and the error variances are estimated separately. Additionally, if the groups are unbalanced in size, HLM improves the estimation of effects of small groups by adjusting them to be similar to the effects for the total sample. That is, each coefficient is weighted proportional to its precision, and for groups with few students, the coefficients for those groups are estimated by "borrowing" information from other data because similar estimates exist for larger groups (Ethington, 1997). As such, HLM improves the estimates in comparison with the disaggregate and aggregate procedures.

Bryk and Raudenbush (1992) illustrated the comparison of three modeling procedures using nested data: HLM, disaggregate, and aggregate analyses. Specifically, they examined the relationship between school characteristics and teachers' sense of efficacy in their work. The comparison of the three methods suggests that in disaggregate analysis the estimated coefficients are robust and generally similar to what would be obtained by using HLM analysis. The magnitudes of the effects of the same variable on other variables are generally quite similar. But, the effect is more likely to be statistically significant in the individual-level analysis because the standard errors are underestimated due to the dependence among individuals within groups. Thus, a researcher is more likely to conclude that an effect is statistically significant when it is actually not. Consequently, incorrect conclusions would be drawn relative to the importance of a variable in predicting the dependent variable in an disaggregate analysis. Using aggregate analysis, the coefficients are not robust, so researchers cannot trust the parameter estimates. Hierarchical analyses, on the other hand, provides the most accurate estimates of the coefficients and the best estimates of the standard errors regardless of the degree of dependency within groups and small imbalances between groups.

Nevertheless, the most dramatic difference between the three methods is in the estimates of the proportion of variance explained in the data. In Bryk and Raudenbush’s (1992) examples using individual-level analyses, only 5.4 % of the variance in the dependent variable was explained by the school characteristics because, at that level, the combined between- and within-group variability was used as the denominator in the calculation. Using the aggregate analysis, the proportion explained in the same dependent variable was 42.6 % because the variability among individuals was lost and only the between-group variability was assessed. In hierarchical analysis, however, the proportion of variance explained in the same dependent variable was 63.1%. Thus, HLM provides the best estimate of the explained variance because, unlike the disaggregate analysis, it is not affected by the dependencies among individuals within groups. And, unlike the aggregate analysis, HLM is not affected by the unreliability caused by the aggregation of data. All of the advantages of HLM over the other alternative analyses dealing with nested data are well supported by simulation studies (see, for example, Barcikowski, 1981; Busing, 1993; Goldstein, 1995; Goldstein & Rasbash, 1996; Kim, 1990; Kish, 1965; Kreft, 1996; Rodriguez & Goldman, 1995; Tate & Wongbundhit, 1983; Van der Leeden & Busing, 1994).

Of course, multilevel modeling is also subject to a number of assumptions that must be carefully evaluated. Multilevel analysis must assume the stability of exposure and covariate distributions over time to ensure that the sample distributions are representative of the population distributions. This assumption will be suspect, of course, when individuals change their behaviors or drop out of a study while it is still progressing. HLM analysis can also suffer from the same problems as in aggregate analysis because poor contextual measures can be constructed from arbitrary administrative or political units (Greenland, 2001).

The accuracy of multilevel analysis also depends on a few assumptions: 1) the dependent variable is assumed to be normally distributed; 2) the independent variables are assumed to be not too highly correlated, which means that collinearity is assumed to be low; and 3) all theoretically defensible levels are assumed to be included in the model as omitting important levels will cause the model to be misspecified. If this happens, incorrect generalizations will be made, and any consequent policy may direct resources to solve a problem at the wrong level. For example, if the classroom level, more than the school level, is the level where the instructional resources would affect the students' achievement, then resources should be directed to the classroom level instead of to the school level. However, this would not be observed if only two levels, the student and the school, are used in an HLM analysis.

Multilevel analysis also faces a major limitation by requiring very large data sets that include many individuals and a number of groups. There are many potential confounding factors, and for multilevel model to reliably control for some of these confounding factors, according to the rule of thumb, there should be at least 30 groups and 30 individuals per group to provide meaningful effect parameters for both individuals and groups. In school effectiveness research, because one of the interests is the group effects, there should be at least 50 groups (Hox, 2002). Regardless of whether these assumptions are reasonable, HLM is more appropriate than either individual analysis that disaggregate groups or aggregate analysis that aggregate individuals into groups.

A Typology for Choosing Modeling Techniques

Table 3 presents a typology of research designs and analyses including multilevel analysis. It identifies four possible sets of situations depending on the type of data that are available in data sets (Subramanian, Jones, Kaddour, & Krieger, 2009).

Table 3

A Typology of Research Designs

	Independent variables	
Dependent variable	x (individual-level)	X (group-level)
y (individual-level)	[y, x] Traditional OLS analyses	[y, X] Multilevel analysis
Y (group-level)	[Y, x] ______________	[Y, X] Ecological analysis

In this table, [y, x] represents a typical individual-level OLS analysis, within which the measurements of the independent variables [x] and dependent variables [y] are both measured at the individual-level, and the contextual effect is not considered. The [Y, X] design represents the situation where the independent and dependent variables are both measured at the group-level. As such, the ecological analysis is most appropriate for data at the group level. As discussed previously, even though the results are not intended to be inferred to individuals, the analysis may be biased because of uncontrolled confounding factors. Of course, it may be even more biased if the contextual effects are assumed to equally affect all the individuals. The [Y, x] represents the research design where the dependent variable is measured at the group-level and the independent variables are measured at the individual-level. This design is not generally of interest to researchers in the social sciences because they are more interested in how group patterns influence individual behaviors instead of how the behaviours of individual influence group patterns.

Finally, the [y, X] is the multilevel design, in which an ecologically measured independent variable [X] is linked to an individual-level dependent variable [y]. Multilevel analysis, such as HLM, can be used for this design which considers individual and contextual effects simultaneously, allowing the effects of both individuals and groups to be assessed. Researchers must, however, be careful to create theoretically meaningful group-level variables and defensible models for the various levels. To adequately analyze school differences, Aitkin and Longford (1986), for example, stated that the minimum requirements are student-level data including the dependent variables and relevant background variables, together with relevant school and district variables, along with explicit modeling of the multilevel structure and a careful analysis of the interactions between these explanatory variables at different levels. In essence though, multilevel modeling has considerable advantages over both disaggregate and aggregate analyses, the modeling must be based on strong theoretical arguments. These arguments for analysis in this doctorate dissertation are presented in Chapter 3.

Summary

School effectiveness is multilevel because of the nested nature of schooling, within which students are grouped within classrooms or by teachers; classrooms or teachers are grouped within schools; schools are grouped in school districts, which are grouped in provinces. Through direct and indirect cross-level effects, these hierarchical layers in the schooling system eventually affect the students' learning opportunities. At the same time, given the same opportunities for learning, students respond differently. Often, this cross-level effectiveness is modeled in two ways, disaggregate and aggregate analyses, both of which evaluate the cross-level impact at a single level. Unfortunately, both procedures are flawed, especially when data are aggregated because researchers often assume that the relationships at the group level are the same at the individual level. Unfortunately, this assumption is usually not true. Ignoring the disparity between the ecological-level and the individual-level relationships often results in researchers committing an ecological fallacy.

Conceptually, an ecological fallacy refers to the incorrect assumption that the relationships between variables observed at the aggregate, or ecological, level are necessarily the same at the individual level. This implies that the pattern of correlations displayed at the higher, or the ecological, level cannot be simply generalized to individuals. Normally, the ecological coefficients are much higher than the individual coefficients and are often not realistic. The direct cause of this disparity is that the ecological coefficients cannot link an individual's outcome to this individual's exposure to the context, nor can it link one individual variable to other individual variables. This failure makes it almost impossible to infer the relationships at the individual level from the ecological coefficients. Unidentifiable and uncontrollable confounding factors are the main reason for this broken linkage between variables at the two levels.

Given these reasons, multilevel modeling, such as HLM, is a more appropriate tool to analyze the cross-level effects in nested educational data. Specifically, in school effectiveness research, researchers are interested in how the organizational structure of schools influences the students' learning over and above the influence of the students' own background, attitudes, and aptitudes. Multilevel modeling can appropriately address this issue because it adequately conceptualizes the effects of schooling by not only modeling how resources are differentially allocated to provide learning opportunities for students, but also by modeling how students differentially respond to the available learning opportunities. Nevertheless, the modeling of multilevel effects should be based on strong theoretical reasoning so that important group effects and theoretically defensible levels are specified. Moreover, the purported group effects should be measured directly rather than indirectly whenever possible.

Multilevel models, or more specifically Hierarchical Linear Models, are composed of sequential linear regression equations, which have different sub-models depending on the research questions. In this dissertation, I use two-level models with schools, teachers, and classrooms at one level and students at a lower level due to the availability of data in the TIMSS study. Starting with one-way ANOVA with a random-intercept and using this as the baseline model, I entered the school-level variables and student-level variables into one-way ANCOVA or full models with fixed coefficients. The data that I used and how the variables were constructed from the data are presented in the next chapter.

CHAPTER 3

METHODOLOGY

Chapter 3 describes the methodology used in this dissertation. First, the Trends in International Mathematics and Science Study (TIMSS) 2007 is introduced including its sampling procedure, the achievement tests and questionnaires, and the Canadian sample. Second, the variables used in this dissertation are introduced including their definitions, the reason for including them, and the way they were measured. As introduced in Chapter 1, the students' mathematics score was the dependent variable. The independent variables were organized into four groups, the provincial variable, the school variables, the teacher and classroom variables, and the student variables. More specifically, the differences between British Columbia and Ontario were evaluated by analyzing the data in each of the two provinces separately. The group of school variables included four variables (the school's physical resources, instructional resources, average SES of students, and administration). There were eight teacher and classroom variables (the teacher's gender, experience, instructional time, assignment of homework, attitudes towards teaching, the homogeneity of students in classroom, and the physical and instructional resources available in classroom). Additionally, there were five student variables (the student's gender, SES, instrumental motivation, educational expectations, and effort in completing homework). Finally, from the review of literature and the descriptive statistics presented, four hypotheses that were tested in this study are listed.

The Sample

The sample of students selected for this dissertation was taken from the Canadian TIMSS 2007 data. This TIMSS study was designed to assess the performances of students in mathematics and science at grade four and grade eight in a number of countries and was designed to inform public institutions, parents, and other citizens about the effectiveness of the curricula and instructional methods, and to help administrators, principals, and teachers improve their teaching and their students' learning of mathematics and science. As a result, the TIMSS study includes a large number of students as well as their teachers and schools. TIMSS 2007 was the fourth cycle in this ongoing study following earlier cycles in 1995, 1999, and 2000. Specifically, in 2007, a total of 183,150 students from 37 countries and seven regional entities participated in the fourth grade study, and a total of 241,613 students from 50 countries and seven regional entities participated in the eighth grade study. The large sample for each country was selected through the carefully designed sampling procedure in TIMSS.

Sampling Procedures in TIMSS

As noted, the target populations in TIMSS are all fourth and eighth graders in each of the participating countries and regional entities. The sampling frame was designed to ensure that the data from students provide accurate estimates of the national or regional populations of students in each of the two grades. To obtain representative samples, in each TIMSS study, a two-stage sampling procedure was used. In the first stage, a random sample of schools was selected with a probability proportional to the school size; and in the second stage, one or two intact fourth and/or eighth grade classes of students (usually mathematics classes) were sampled. Following this sampling procedure, about half of the countries and regional entities sampled classes from at least 150 schools in each of grades four and eight, which is the minimum number required to meet the TIMSS sampling requirement. For each grade, most countries sampled either one or two classrooms of students from each sampled school. Consequently, in each country, between 4,500 and 6,000 students, enrolled in at least 150 schools, were included. This is a very effective and efficient sampling frame, which becomes more evident in chapters 4 and 5. However, the resulting samples have a complex structure that must be considered when the data are analyzed. In particular, sampling weights need to be applied or re-sampling techniques, such as a jackknife procedure, need to be used to correctly estimate the sampling variances (Foy & Olson, 2009).

To obtain representative samples, the TIMSS researchers also clearly documented the schools and students that had been excluded from the sampling frame. Exclusions could occur for specific students, classrooms, and schools. Specifically, countries could exclude schools that: 1) were geographically inaccessible; 2) were extremely small; 3) had curricula or school structures that were radically different from the mainstream education system; and 4) only provided instruction to students in an excluded category, such as schools for the blind or the deaf. In addition, each country adapted two other exclusion rules when creating their sample of students. First, they excluded students with intellectual disabilities and emotional or mental problems that made it impossible for them to understand the instructions on the test; and second, they excluded physically disabled students who could not perform in the TIMSS testing situation, for example, blind, deaf, and students who were non-native-language speakers and were unable to read the national language used in the test. However, students with minor functional disabilities who were able to respond, however, were included in the testing. These two exclusion rules ensured that students were not excluded solely because of their poor academic performances or because of normal disciplinary problems. As a consequence, in most countries and regional entities, the percentage of excluded students was less than 5%.

Administration of TIMSS Tests

The TIMSS 2007 assessment instrument contained 353 items at the fourth grade (179 items in mathematics and 174 items in science) and 429 items at the eighth grade (215 items in mathematics and 214 items in science). At each grade, the items were organized into 14 blocks of mathematics items and 14 blocks of science items. The blocks of test items were then assembled into 14 booklets, each containing 2 blocks of mathematics items and 2 blocks of science items. According to a very carefully designed protocol, all test items were equally represented in the 14 booklets. Each student was required to finish one test booklet. The time normally needed for students to complete a block of items was 18 minutes at the fourth grade and 22.5 minutes at the eighth grade. Consequently, it normally took fourth grade students 72 minutes and eighth grade students 90 minutes to finish the test.

Along with the students' tests in mathematics and sciences, the students, their mathematics and sciences teachers, and their school principals were asked to answer questionnaires in order for the researchers to obtain information on the home and school conditions of the students. In brief, students answered questions about their social backgrounds, psychological dispositions, and their home and school environments; the teachers responded to questions about their backgrounds and attitudes, the characteristics of the physical and social environments in their classrooms, their instructional activities, and the topics they covered. The principals, in turn, responded to questions about the enrolment of students, the school's physical and social characteristics, organization, staffing, and resources.

To ensure that all the students had been exposed to their learning material long enough, the tests and questionnaires for the study were administered near the end of the school year. In particular, countries in the southern hemisphere, where the school year typically ends in November or December, conducted the assessment in October or November, 2006. In countries in the northern hemisphere, where the school year typically ends in June, the assessment was conducted in April, May, or June, 2007.

Canadian TIMSS Sample

Four Canadian provinces, British Columbia, Alberta, Ontario, and Quebec, participated in TIMSS 2007. However, Alberta only administered the assessment to fourth grade students. In this dissertation, grade eight students from two provinces, British Columbia (BC) and Ontario (ON), were chosen because these two provinces mainly offer English programs and they both administered the assessment to eighth grade students. Additionally, only the mathematics scores were used to evaluate the students' performance.

Table 4 reports the populations and the samples of schools, teachers, and students in the TIMSS 2007 study in British Columbia and Ontario. In British Columbia, there were 51,804 grade eight students studying in 433 schools. The TIMSS 2007 program sampled 4,256 students nested in 150 schools, which represented 41,735 students from the provincial student population (80.6%). These students were tested in April (97.3%) and May (2.7%). At the time of testing, the students' average age was 14.06 years. Among these students, 50.4% were girls and 49.6% were boys. Similarly, in Ontario, there were 159,230 grade eight students studying in 2,854 schools. The TIMSS program sampled 3,448 students in 176 schools, which represented 143,755 students from the provincial student population (90.28%). All of these students were tested in April. At the time of testing, the students' average age was 14.03 years. Among these students, 50.6% were girls and 49.4% were boys. All mathematics teachers for these students in both British Columbia and Ontario were invited to participate in the TIMSS survey. Eventually, 185 mathematics teachers in British Columbia and 206 mathematics teachers in Ontario participated.

Table 4

Populations and Samples in British Columbia and Ontario in TIMSS 2007 Data

	British Columbia				Ontario			
	Population	Sample	Obtained	Obtained %	Population	Sample	Obtained	Obtained %
Schools	433	150	134	89.33%	2,854	176	169	96.02%
Students	51,804	4,256	2,392	56.20%	159,230	3,448	2,080	62.37%
Teachers	-	185	167	90.27%	-	206	198	96.11%

In the school data, 16 schools and their 396 students in British Columbia, and seven schools and their 113 students in Ontario were not included because of missing values. How these schools were different from the schools that were included was assessed and the result is reported in Chapter 4 and discussed in Chapter 5. These missing schools were deleted from the analyses to decrease the errors which would have resulted from the replacement of missing values.

Table 5 presents the characteristics of the schools included in this dissertation. Overall, schools in British Columbia had from 24 to 2,090 students, and schools in Ontario had from 39 to 1,367 students. On average, schools in British Columbia were much larger (an average of 883 students) than schools in Ontario (an average of 410 students). In addition, there were more students who were native English speakers in schools in British Columbia than in Ontario. However, compared to schools in British Columbia, schools in Ontario had, on average, five more instructional days per year. Thus, students in Ontario probably spent more time on their school work.

Table 5

Characteristics of Schools in British Columbia and Ontario

	Range		Mean		SD	
	BC	ON	BC	ON	BC	ON
School enrollment	24–2090	39–1367	883.62	410.33	447.28	230.89
School enrollment of 8th grade	3–400	2–307	187.14	72.06	83.31	64.52
Instructional days per year	150–207	168–213	186.68	191.53	7.88	5.00

In the student data, the students' parental educations had considerable missing values, as high as 39%, which was probably caused by a number of students responding that they did not know their parents' educational levels. In both British Columbia and Ontario, students who had missing information about their mothers' education were also likely to have missing information on other important background variables, such as their gender. Dropping these students from analysis has dramatically decreased the amount of missing values for most student variables to below 10% except that the amount of missing values for the father's education was still 19%. For this reason, only the mother's education was included in the analyses. Again, how the students that were dropped were different from the students that remained was assessed. The result is reported in Chapter 4 and discussed in Chapter 5.

As reported in Table 4, 2,392 students in British Columbia and 2,080 students in Ontario have been selected for the analyses. Similarly, the students were around 14 years old at the time of testing and the numbers of male and female students were about the same in both provinces. Additionally, over 75% of these students always or almost always spoke English at home. However, there were obvious differences between the students in British Columbia and Ontario. On average, students in Ontario achieved higher scores in mathematics than students in British Columbia. Also, the students' parents in Ontario had, on average, higher educational levels than the parents of students in British Columbia.

Because the student' mathematics score was used as the dependent variable, only the students' mathematics teachers were selected for analysis. After dropping teachers in the missing schools, 167 teachers in British Columbia and 198 teachers in Ontario were included. In both provinces, about 65% of the teachers were between 30 and 49 years of age. Additionally, almost all the teachers had university degrees and teaching licenses or certificates. However, there were differences between teachers in British Columbia and Ontario. Specifically, in British Columbia, 43% were female teachers and 57% were male teachers and, on average, they had been teaching for about 12.5 years, while in Ontario, 58% were female teachers and only 42% were male teachers; on average, they had been teaching for about 10 years.

Eventually, 134 schools with 167 mathematics teachers and 2,392 students from British Columbia and 169 schools with 198 mathematics teachers and 2,080 students from Ontario were included in the analyses. On average, there were 17.85 students in each selected school in British Columbia and 12.31 students in each selected school in Ontario. There were important differences between the samples included and dropped from analysis. Detailed assessments are reported in Chapter 4 and discussed in Chapter 5. The next section of this chapter introduces the variables analyzed in the dissertation. In total, there was one dependent variable and seventeen independent variables of three groups: four school variables, eight teacher and classroom variables, and five student variables.

Dependent Variable

The dependent variable for this study was the students' academic achievement in mathematics, which was measured by five plausible values in mathematics for each student. Plausible values are defined as random values from the posterior distributions, which can be understood as representing the range of achievement scores that a student might reasonably have achieved if that student had repetitively taken all the test items in the mathematics pool of items. Consequently, the use of plausible values minimizes the measurement error associated with each individually estimated achievement score.

Plausible values have been used in TIMSS as well as in other large-scale national and international assessment studies, such as PISA. To adequately explain the differences in achievement between countries, between schools, and between students, many test items were developed for TIMSS. However, only some of the items were included in the test for each student because students were assigned to answer tests drawn from subsets of an item pool to overcome the conflicting demands of limited testing time of students and a broad coverage of the assessment domain. Consequently, only sub-samples of students responded to each item. To ensure that the student's scores obtained from different test items were comparable, TIMSS 2007 used Item Response Theory (IRT) scaling to improve measures, which used multiple imputation–or plausible values–methodology to obtain reasonably accurate achievement scores for all students. As a result, each student's record in the TIMSS 2007 data set contained five imputed scores in mathematics. Since each imputed score was a prediction based on the limited information in the set of items the student answered, it included a small amount of error. To allow analysts to incorporate this error into analysis, the TIMSS 2007 data provided five separate plausible values for achievement in mathematics. The analysis to infer population scores should be replicated five times, each using a different plausible value. The population estimate is then the average of the five analyses. It is obvious that the standard errors in the combined results include both sampling and imputation errors. Consequently, the errors of individual students are reduced, and more importantly, they are reduced in making inferences from the sample to the population, which is basic information for testing the statistical significance of effect parameters.

Table 6

Descriptive Statistics for the Dependent Variable

Plausible value	Range		Mean		S.D.		t-test
	BC	ON	BC	ON	BC	ON	
MathPV1	265.72–757.33	306.15–732.96	515.51	524.21	69.82	66.08	4.26***
MathPV2	294.81–734.79	306.70–738.62	515.36	524.35	71.35	66.60	4.34***
MathPV3	268.11–733.90	289.59–727.51	515.65	524.36	71.79	66.46	4.19***
MathPV4	252.74–721.25	261.32–734.31	515.95	524.06	70.41	66.94	3.93***
MathPV5	281.91–760.29	308.98–736.78	516.22	524.70	70.19	66.50	4.13***

MathPV, plausible values in the mathematics test.
***$p \leq .001$.

The descriptive statistics for the five plausible values of mathematics and the t-tests of the differences between the means in British Columbia and Ontario are reported in Table 6. As indicated in the procedure used to produce these plausible values, none of them had missing values and every value was normally distributed. As such, the missing values, skewness, and kurtosis are not reported in this table. Nevertheless, this table shows that students in Ontario achieved, on average, around 9 points, or about 13% of a standard deviation, higher than students in British Columbia, which was statistically significant ($p \leq .001$).

Independent Variables

Four groups of independent variables were used in this dissertation, the provincial variable, the school variables, the teacher and classroom variables, and the student variables. This section introduces these independent variables' definitions, the reason for including them, the way they were measured, and their descriptive statistics. Several variables were measured with multiple items. The original items for each of these variables were first assessed by using exploratory factor analysis. When it was clear that the items were measuring the same underlying factor, they were loaded on a single factor using confirmatory factor analysis. The results of these confirmatory analyses are reported in this section as well as their alpha reliability coefficients. For parsimony, the multiple items were simply summed to build each variable.

Provincial Variable

The provincial impact was evaluated by analyzing the data in each of the two provinces. Earlier, it was noted that substantial provincial differences exist in the students' academic achievement so that students in British Columbia generally are thought to have higher academic achievement than students in Ontario. However, recent PISA studies suggest that the differences are changing. Specifically, students in Ontario are doing better.

School Variables

Four independent variables measured at the school level, the school's physical resources, instructional resources, average SES of students, and administration, were included in the analyses. All these variables were obtained from the principals' questionnaire. The physical resources and instructional resources were measured by multiple items.

Physical resources. Physical resources refer to the schools' facilities including the buildings, both internal and external conditions, and the space they have for instructional purposes. Recently, physical resources of schools have become a highly contested area of research, within which school facilities, in general, and certain kinds of facility, in particular, have been studied. Overall, inconsistencies and loosely stated conclusions are evident in this research literature. Nevertheless, a growing body of research has indicated the positive effect of the schools' physical resources by linking the school building quality to the students' academic achievement and attitudes, as well as to the teachers' attitudes and behaviors. More specifically, the quality of the buildings that relates to the human comfort of the teachers and students has been shown to be related to the students' academic achievement. For example, the age of the building (Earthman & Lemasters, 1996; McGuffey & Brown, 1978), climate control (Cash, 1993; Earthman, 2004; Tanner, 2000; Tanner & Lackney, 2006), lighting (Heschong Mahone Group, 1999), and acoustical control (Cash, 1993; Earthman, 2004), have all been reported to be relevant to the students' academic achievement. These studies suggest that school buildings that are in poor shape lead to students' reduced learning, but this literature lacks a logical and coherent theoretical argument explaining why the facilities are causally linked to the students' educational achievement.

Other researchers argue that there is very little empirical evidence supporting this commonly held belief that high-quality school facilities have a positive impact on the students' academic achievement (Picus, Marion, Calvo & Glenn, 2005). Burtless (1996) found that only half of the articles he reviewed provided evidence that increasing the expenditures on physical improvements to schools and classrooms resulted in improved academic achievement of students. It is difficult to determine any definite and consistent findings from this large number of studies. In many cases, the researchers could not explain why the results obviously did not support the common sense notion that the better quality of schools increases the achievement of students. For example, Cash (1993) found that higher rated schools unexpectedly had more disciplinary problems than lower-rated schools. Similarly, Earthman, Cash, and Van Berkum (1995) found that when considering the structural factors alone, students in substandard buildings outscored those in above-standard buildings. In a number of other empirical studies, the relationship between the physical resources of schools and the students' achievement was weak (Cervini, 2009; Dearden et al., 2002**;** Wößmann, 2003).

It is possible that the inconsistencies in the studies may result from small distributional ranges in the variables measuring the school facilities because, at least in an industrialized country, Canada, specifically, school facilities do not vary substantially. Also, when examining how the building environment affects the students' learning, ideally researchers would control a number of confounding variables, such as the students' parental education, parental occupation, family income, family structure, the students' effort in studying, the teachers' competency and attitudes, etc. (Ma, Ma, & Bradley, 2008). Unfortunately, few studies have reported data on these confounding variables, and, therefore, they were not controlled (Picus, Marion, Calvo, & Glenn, 2005). Without controlling for important confounding variables, it is impossible to draw definitive conclusions about the effects of the schools' facilities on the students' learning.

Some other researchers argue that the conventional measurements of school facilities using an engineering perspective, such as the Facility Condition Index (FCI)

(Organization for Economic Cooperation & Development, 2000; Piper, 2004), does not take into account the purpose of schools. Without identifying the purpose of a school's facility, it is impossible to assess how good the facility is for teaching and learning. Consequently, analyses based on engineering instruments often produce vague results (Roberts, 2009). In comparison, studies that assessed the principals' perspective on the school facilities that were relevant to teaching and learning found meaningful connections between the facility conditions and the educational outcomes (Bullock, 2007; Earthman, Cash, & Van Berkum, 1995; Roberts, 2009). In general, measures using an educator's assessment of the schools' facilities seem to produce more consistent and meaningful results.

Nevertheless, these inconsistent findings seem largely to result from the aggregation analysis commonly used in studies of the physical resources in which the student variables were aggregated to the school level and then the school was used as the unit of analysis. In a recent summary of 20 published studies about the school facilities by 21st Century School Fund (2009), 14 have linked school facilities to the aggregated test scores of the students at the school level (Blincoe, 2008; Boese & Shaw, 2005; Buckley, Schneider, & Shang, 2004; Bullock, 2007; Cellini, Ferreira, & Rothstein, 2008; Crampton, 2009; Duran-Narucki, 2008; Hughes, 2006; Lewis, 2000; Picus, Marion, Calvo, & Glenn, 2005; Schneider, 2003; Sheets, 2009; Stevenson, 2001; Tanner, 2009). However, the procedure of aggregating data to the school-level reduces the variance in the independent and dependent variables as well as decreasing the units of analysis to the number of schools which often results in biased, most often overestimated, coefficients (Hanushek et al., 1998; Raudenbush & Wilms, 1995; Robinson, 1950). Moreover, it is a serious bias to assume that the relationship among variables at the school level would be the same as the relationship among the same variables at the student level. As discussed previously in Chapter 2, the inconsistent findings in the current literature may result from the aggregation bias in many studies.

In this dissertation, the physical resources were obtained from three items in the principals' questionnaire: "Is your school's capacity to provide instruction affected by a shortage or inadequacy of any of the following?" The responses were made on a 4-point scale.

None	A little	Some	A lot
1	2	3	4

1. School buildings and grounds.
2. Heating/cooling and lighting systems.
3. Instructional space (e.g., classrooms).

The responses to these items were reverse scored so that higher values indicated more physical resources in the schools. The correlation coefficients, factor loadings, percentages of variance explained, and reliability coefficients of these items for schools in British Columbia and Ontario are reported in Table 7. The correlation coefficients were high, ranging from .38 to .63 in British Columbia and .49 to .65 in Ontario. Also, the high and consistent factor loadings on a single factor confirmed the assumption that the items measured the same construct in both provinces. In fact, one factor explained 67.52 % and 72.03% of the variance in the data from British Columbia and Ontario, respectively. In addition, these items were highly reliable as indicated by the Cronbach's Alpha coefficients, .78 in British Columbia and .81 in Ontario.

Table 7

Correlation and Factor Analyses of Items for the Physical Resources Variable

Correlations					Factor loadings	
Items	1	2	3		BC	ON
1		.59**	.63**		.90	.90
2	.63**		.38**		.78	.82
3	.65**	.49**			.81	.84
				Percentage of variance explained	69.17%	72.66%
				Cronbach's Alpha	.78	.81

The British Columbia correlation matrix is above the diagonal, and the Ontario matrix is below the diagonal.
**$p \leq .01$.

Table 8

Descriptive Statistics for the Physical Resources Variable

Province	Range	Mean	SD	Skewness	Kurtosis	Missing	t-test
British Columbia	5–12	10.23	2.07	-1.01	-.19	.7%	.16
Ontario	3–12	10.19	2.19	-1.50	1.94	1.2%	

The three items were summed to create the physical resources variable. Table 8 reports the descriptive statistics for this variable in British Columbia and Ontario, and the t-test for the difference between means in the two provinces. As expected, this variable was roughly normally distributed in both provinces and it was slightly skewed to the left meaning that most schools' facilities were in good condition. While the distribution was slightly platykurtic in British Columbia, it was slightly leptokurtic in Ontario meaning that in Ontario physical resources were more evenly distributed among schools. Though Ontario seemed to have less developed physical resources, the difference may be due to random variation ($t=.16$, $p>.05$).

Instructional resources. The availability of instructional resources in schools is one of the few factors that are assumed to affect the students' learning and over which schools have almost complete control. As such, researchers have sought to predict the student scores on standardized tests, as a function of the instructional resources in schools and the students' demographic characteristics (Picus, 2001). Like the research on the schools' physical resources, the research findings have been inconsistent; some studies have shown that the schools' instructional resources had little or no effect on the students' achievement (Coleman et al., 1966; Hanushek, 1986, 1996, 2003; Jencks, Smith, Ackland, Bane, Cohen, Ginter, Heyns, & Michelson, 1972), while other studies have shown that they had small positive effects (Hedges, Laine, & Greenwald, 1994; Krueger, 1999; Verstegen & King, 1998). These inconsistent results may also be a result of weaknesses in research methodology, specifically, the restriction of variance for the instructional resources, not controlling for confounding variables, and the aggregation bias as previously discussed in Chapter 2.

In this dissertation, the schools' instructional resources were measured by six items that were included in the principals' questionnaire: "Is your school's capacity to provide instruction affected by a shortage or inadequacy of any of the following?" The responses were made on a 4-point scale.

None	A little	Some	A lot
1	2	3	4

1. Instructional materials (e.g., textbook).
2. Budget for supplies (e.g., paper, pencils).
3. Computers for mathematics instruction.
4. Computer software for mathematics instruction.
5. Library materials relevant to mathematics instruction.
6. Audio-visual resources for mathematics instruction.

The responses to these items were reverse scored so that higher values indicated more available instructional resources in schools. The correlation coefficients, factor loadings, percentages of variance explained, and reliability coefficients of these items in British Columbia and Ontario are reported in Table 9. The correlation coefficients were relatively high in both provinces. However, they seemed to be slightly higher and more consistent in British Columbia, where they ranged from .46 to .89, than in Ontario, where they ranged from .23 to .77. Nevertheless, the high and consistent factor loadings of these items on a single factor confirmed the assumption that they measured the same construct in both provinces. This factor explained 66.35% and 54.51% of the variance for schools in British Columbia and Ontario. Moreover, they were highly reliable as indicated by the Cronbach's Alpha coefficients, .90 in British Columbia and .83 in Ontario.

Table 9

Correlation and Factor Analyses of Items for the Instructional Resources Variable

Correlations							Factor loadings	
Items	1	2	3	4	5	6	BC	ON
1		.69**	.58**	.51**	.53**	.46**	.76	.61
2	.60**		.52**	.50**	.51**	.47**	.75	.62
3	.23**	.26**		.89**	.61**	.56**	.87	.69
4	.28**	.33**	.74**		.63**	.61**	.86	.83
5	.44**	.41**	.37**	.54**		.76**	.83	.82
6	.32**	.34**	.46**	.68**	.77**		.81	.84
				Percentage of variance explained			66.35%	54.51%
					Cronbach's Alpha		.90	.83

The British Columbia correlation matrix is above the diagonal, and the Ontario matrix is below the diagonal.
**$p \leq .01$.

The six items were summed to create the schools' instructional resources variable. The descriptive statistics for this variable in British Columbia and Ontario, and the t-test for the difference between means in the two provinces are reported in Table 10. The distribution of this variable was slightly skewed to the left in both provinces meaning that the principals indicated that most schools' instructional resources were well supplied. While the variable was slightly leptokurtic in British Columbia, it was slightly platykurtic in Ontario. It is worth noting that British Columbia schools had significantly more instructional resources than Ontario schools as judged by the principals ($t=2.25$, $p \leq .05$).

Table 10

Descriptive Statistics for the Instructional Resources Variable

Province	Range	Mean	SD	Skewness	Kurtosis	Missing	t-test
British Columbia	6–24	19.33	4.32	-.92	.10	3.7%	2.25*
Ontario	8–24	18.25	3.97	-.40	-.56	1.2%	

*$p \leq .05$.

Average socioeconomic status (SES) of students. The average SES of students in schools is a prominent concern in the research literature. Reference group theory, in particular, has been used to draw attention to the values and goal-orientations of students in schools, and has been seen as characterizing the educational environments created by the students (Bidwell, 1972; Haller & Woelfel, 1972; Woelfel & Haller, 1971). Substantial research literature shows that students attending schools with a higher percentage of high status students have an advantage that would not otherwise be expected on the basis of their own status origins (Alexander & Eckland, 1975; Alwin & Otto, 1977; Bain & Anderson, 1974; Bidwell, 1972; Haller & Woelfel, 1972; Hauser, Sewell, & Alwin, 1976; Meyer, 1970; Nelson, 1972; Willms & Smith, 2004; Woelfel & Haller, 1971). The accumulated evidence for the relevance of this variable is, in fact, very impressive.

Some researchers have tried to explain the positive impact of a higher average SES of the student body on the students' academic achievement. McDill, Meyers, and Rigsby (1967), McDill and Rigsby (1973), and McDill, Rigsby, and Meyers (1969) suggested that academically oriented normative climates are created in upper- and middle-class schools, and this climate enhances the academic achievement of all students.

Further, the researchers have noted that when individual climate variables are controlled, the average SES levels of schools are virtually unrelated to the educational outcomes of the students. This finding is consistent with an interpretation of the claim that school differences in student-body SES represent differences in the value climate of schools. In other words, through the well-documented processes of reference group influences, the enhanced likelihood of creating close relationships with high SES peers in schools with high average SES levels rather than low levels is thought to yield numerous positive educational benefits for low SES students (Duncan, Haller, & Portes, 1968; Haller & Butterworth, 1960; Herriott, 1963; Kandel & Lesser, 1969; Kelly, 1952; Sewell, Haller, & Ohlendorf, 1970; Simpson, 1962; Woelfel & Haller, 1971). Nevertheless, some studies argue that the average SES of schools has a compositional effect, which is artificial due to statistical reasons, instead of a contextual effect (Alexander, Fennessey, McDill, & D'Amico, 1979; Nash, 2003). Hauser (1970, 1971) contended that some contextual effects may, in fact, be a statistical artifact caused by the underspecified mathematical models.

In this dissertation, the average SES of students was obtained from their principals' responses to a question about the approximate percentage of students in their schools who came from economically disadvantaged homes. The responses the principals could select were: 0 to 10%, coded as 1; 11% to 25%, coded as 2; 26% to 50%, coded as 3; and more than 50%, coded as 4. The responses were reverse coded so that higher scores indicated more students from higher SES homes in the schools. The descriptive statistics for this variable in British Columbia and Ontario, and the t-test for the difference between means in the two provinces are reported in Table 11. In both British

Columbia and Ontario, between 26% and 50% of the students in an average school were from economically disadvantaged homes. The distribution in both provinces was slightly platykurtic and slightly skewed to the left. In addition, Ontario had more missing values, which probably indicated that more principals in Ontario were not sure about the SES of their students.

Table 11

Descriptive Statistics for the Average SES of Students Variable

Province	Range	Mean	SD	Skewness	Kurtosis	Missing	t-test
British Columbia	1–4	3.05	.87	-.51	-.68	4.5%	.83
Ontario	1–4	3.13	.87	-.68	-.40	7.1%	

Administration. One of the fundamental interests in the school improvement research literature concerns the apparent impact of principals on their students' academic achievement. In fact, research findings from a number of countries draw similar conclusions that schools that make a difference in the students' learning are led by effective principals. Even though the leadership of principals is far from simple, much anecdotal evidence suggests that effective school administrators are able to strengthen school programs, create positive learning conditions, improve teachers' morale, and inspire teachers to teach more effectively. Furthermore, effective principals have a strong interest in improving their teachers' instruction, and they can guide teachers to help their students learn the course material (Andrews & Soder, 1987; Bossert, Dwyer, Rowan, & Lee, 1982; Hallinger, 2003; Hallinger & Heck, 1996). In essence, effective principals work hard to maintain a positive school climate, within which both students and teachers strive toward higher levels of academic achievement.

In this dissertation, the administration of schools was assessed by the principals' responses to a question about the approximate percentage of time they spent on administrative duties. The descriptive statistics for this variable in British Columbia and Ontario, and the t-test for the difference between means in the two provinces are reported in Table 12. The distributions in both British Columbia and Ontario were normal. The principals in both provinces spent a considerable portion of their time on administrative duties, but in British Columbia, the principals spent significantly more time ($t=3.36$, $p \leq .01$) than principals in Ontario.

Table 12

Descriptive Statistics for the Administration of Schools Variable

Province	Range	Mean	SD	Skewness	Kurtosis	Missing	t-test
British Columbia	5–90	50.26	18.59	-.18	-.36	5.2%	3.36**
Ontario	5–85	43.07	17.53	.27	-.39	2.4%	

$^{**}p \leq .01$.

Teacher/Classroom Variables

Eight teacher and classroom variables, the teacher's gender, experience, instructional time (measured by three separate variables, instructional time in teaching numbers and algebra, teaching geometry, and teaching data and chance), assignment of homework, attitudes towards teaching, homogeneity of students in classroom, and physical and instructional resources available in classroom, were included in analysis. All these variables have been obtained from the questionnaire answered by the students' mathematics teachers. All teacher/classroom variables except for the teacher's gender and experience were measured by multiple items.

Teacher's gender. The effect of teachers' gender has not been examined systematically in the educational literature. The little research that has been conducted suggests that the influence of the teachers' gender on their students' achievement has had mixed results. Saha (1983) reported that in 21 less-developed countries, males were generally more successful than females in teaching mathematics and science. Consequently, students with male teachers had higher achievement in mathematics and science than students with female teachers (Carnoy, 1971; Klees, 1974; Warwich & Jatoi, 1994). In contrast, two other studies found that students with female teachers had significantly higher achievement scores in mathematics (Mwamwenda & Mwamwenda, 1989) and reading (Lam, Tse, Lam, & Loh, 2010) than students with male teachers. Additionally, some other studies showed that the teachers' gender had no impact on their students' achievement (Andrew & Herb, 2005; Bloom, 2008; Holmlund & Sund, 2005; Wiles 1992). Even though the impact of the teachers' gender on their students' mathematics achievement is not clear, it is evident that there are significant differences in the preferred teaching techniques used by male and female teachers (Lam, Tse, Lam, & Loh, 2010). This evidence suggests that the teachers' gender, as an independent variable, may have an effect on the students' mathematics achievement.

In this dissertation, male mathematics teachers were coded "0" and females were coded "1". The percentage of female teachers is reported in Table 13. As indicated by the chi-square test, also reported in this table, there were significantly more female mathematics teachers in Ontario than in British Columbia. It is noticeable that 6.0% of the responses were missing in British Columbia while only 2.9% were missing in Ontario.

Table 13

Descriptive Statistics for the Teacher's Gender Variable

Province	Percentage of Females	Missing	χ^2
British Columbia	43%	6.0%	7.27***
Ontario	58%	2.9%	

***$p \leq .001$.

Teacher's experience. The years that teachers have been teaching indicates the experience they have had. When we think of highly capable teachers, first we think about the adequacy of their education, and second we think about their teaching experience. In Canada, since most teachers have at least university degrees, the experience then seems more important. As commonly perceived, the more years of teaching, the more proficient teachers become, and the more capable they are at handling daily routines and difficult situations. As such, the evaluation of teacher quality is mostly related to their educational level and years of teaching experience. In fact, teachers in North America receive increased salaries as they gain teaching experience, which recognizes the importance of their experience. Some studies have, in fact, found teachers' effectiveness increases with their years of experience (Klitgaard & Hall, 1974; Murnane & Phillips, 1981), but surprisingly other studies have found that the teachers' education and experience have no consistent effects on their students' academic achievement (Hanushek, 1986, 1996, 1997; Palardy & Rumberger, 2008).

It is worth noticing that in the studies that reported positive effects, the benefits of experience appear to level off at about five years (Rivkin, Hanushek, & Kain, 2005; Rockoff, 2004; Rosenholtz & Bassler, 1986). The benefits of experience may interact with the opportunities provided in their schools. Veteran teachers who emphasize continual learning may continue to improve their teaching performances, which may affect their students' academic achievement (Rosenholtz & Smylie, 1984). Of course, some very well-prepared beginning teachers can also be highly effective in helping their students learn the subject matter (Andrew & Schwab, 1995; Denton & Peters, 1988). It is possible that the uneven effects of experience in these cross-sectional studies could be the result of cohort effects. For example, some researchers have suggested that cohorts of teachers that were hired in times of shortage may not do as well as those hired when school administrators could be more selective, or at various times there may be a disproportionate attrition of more capable teachers (Murnane & Phillips, 1981; Vance & Schlechty, 1982). Nevertheless, like other professions, teachers with longer teaching experience are likely to help their students learn better.

In this analysis, the years that the teachers had been teaching were coded into an interval variable as follows:

1 to 4 years	1	5 to 9 years	2	10 to 14 years	3
15 to 19 years	4	20 to 24 years	5	25 to 29 years	6
30 and above	7				

The descriptive statistics for the teacher's experience variable in British Columbia and Ontario, and the t-test for the difference between means in the two provinces, are reported in Table 14. As indicated, the distribution was slightly skewed to the right especially in Ontario. While the distribution was slightly platykurtic in British Columbia, it was slightly leptokurtic in Ontario. On average, mathematics teachers in British Columbia had significantly more years of teaching experience than teachers in Ontario (t=2.39, $p \leq .05$). Similar to gender, this variable had more missing values (9.9%) in British Columbia than in Ontario (5.4%).

Table 14

Descriptive Statistics for the Teacher's Experience Variable

Province	Range	Mean	SD	Skewness	Kurtosis	Missing	t-test
British Columbia	1–7	3.09	1.76	.72	-.41	9.9%	2.39*
Ontario	1–7	2.64	1.76	1.26	.83	5.4%	

*$p \leq .05$.

Teacher's instructional time. The instruction delivered by teachers reflects, to a certain degree, the opportunities that students have to learn mathematics. Logically, if students are not exposed to mathematics in schools or the emphasis is not sufficient, they will have little opportunity to learn the subject matter. The amount of instruction can be roughly evaluated by the content covered (Borg, 1979; Good, Grouws, & Beckerman, 1978) or the percentage of test items on a standardized test that have been taught by the teachers (Cooley & Leinhardt, 1980; Dunkin & Doenau, 1980; Wyne & Stuck, 1982). Also, the amount of instruction can be evaluated by the amount of time the teachers devote to instruction. It has been found, for example, that students achieve more in classes where they spend most of their time being taught or supervised by teachers than in classes where they spend more time working on their own or, in some cases, not working at all (Brophy & Evertson, 1976; Good & Becherman,1978). Of course, the amount of time devoted to teaching is bounded by the length of a school day and the number of days in a school year (Brophy & Evertson, 1976; Coker, Medley, & Soar, 1980).

In this analysis, three variables, teaching numbers and algebra, teaching geometry, and teaching data and chance measured the teachers' instructional time. They were all constructed from a number of items in the mathematics teachers' questionnaire that the teachers had three options as follows:

1 Mostly taught before this year;

2 Mostly taught this year (If a topic was taught half this year but not yet completed);

3 Not yet taught or just introduced (if a topic is not in the curriculum).

The responses to these items were reverse scored so that higher values indicated more instructional time spent on teaching the specific topics.

Teaching numbers and algebra. In total, 18 items measured the instructional time in teaching numbers and algebra variable:

n1. Whole numbers including place value, factorization, and the four operations.

n2. Computations, estimations, or approximations involving whole numbers.

n3. Common fractions including equivalent fractions and ordering of fractions.

n4. Decimal including place value, ordering, and converting to common fractions (and vice versa).

n5. Representing decimals and fractions using words, numbers, or models (including number lines).

n6. Computations with fractions.

n7. Computations with decimals.

n8. Representing, comparing, ordering, and computing with integers.

n9. Ratios (equivalence, division of a quantity by a given ratio).

n10. Conversion of percents to fractions or decimals and vice versa.

a1. Numeric, algebraic, and geometric patterns or sequences (extension, missing terms, generalization of patterns).

a2. Sums, products, and powers of expressions containing variables.

a3. Evaluating expressions for given numeric value.

a4. Simplifying or comparing algebraic expressions.

a5. Modeling situations using expressions.

a6. Evaluating functions/formulas for given values of the variables.

a7. Simple linear equations and inequalities, and simultaneous (two variables) equations.

a8. Equivalent representations of functions as ordered pairs, tables, graphs, words, or equations.

The correlation coefficients, factor loadings, percentages of variance explained, and reliability coefficients of these items in British Columbia and Ontario are reported in Table 15. Though the correlations varied, the items loaded highly on a single factor suggesting that they measured the same construct.

Table 15
Correlation and Factor Analyses of Items for the Teaching Number and Algebra Variable

Correlations																			Factor loadings	
Items	n1	n2	n3	n4	n5	n6	n7	n8	n9	n10	a1	a2	a3	a4	a5	a6	a7	a8	BC	ON
n1		.63**	.47**	.34**	.34**	.24**	.46**	.25**	.26**	.29**	.17*	.03	.09	.06	.15	.03	-.12	-.09	.41	.37
n2	.67**		.38**	.37**	.39**	.20**	.37**	.27**	.29**	.27**	.18*	.05	.12	.12	.08	.08	-.04	-.03	.43	.49
n3	.32**	.45**		.61**	.57**	.60**	.53**	.46**	.43**	.49**	.14	.17*	.19*	.19*	.19*	.13	.06	.12	.66	.55
n4	.32**	.37*	.55**		.56**	.40**	.56**	.36**	.30**	.36**	.17*	.18*	.22**	.16*	.17*	.14	.03	.07	.56	.64
n5	.38**	.37**	.59**	.58**		.29**	.46**	.34**	.29**	.27**	.10	.14	.10	.15	.13	.18*	.02	.08	.53	.54
n6	.25**	.29**	.54**	.45**	.46**		.54**	.48**	.54**	.50**	.19*	.29**	.28**	.29**	.32**	.21**	.24**	.19*	.68	.60
n7	.24**	.26**	.39**	.53**	.44**	.40**		.41**	.39**	.37**	.20*	.16*	.18*	.16*	.26**	.17*	.07	.12	.61	.49
n8	.18*	.11	.10	.22**	.22**	.17*	.24**		.44**	.39**	.15	.15*	.14	.13	.13	.09	.08	.03	.52	.43
n9	.14	.16*	.22**	.28**	.28**	.35**	.26**	.39**		.70**	.18*	.17*	.30**	.30**	.28**	.22**	.18*	.25**	.65	.50
n10	.15*	.20**	.45**	.46**	.34**	.30**	.31**	.21**	.51**		.18*	.12	.22**	.22**	.23**	.20*	.13	.16*	.60	.50
a1	.21**	.25**	.16*	.28**	.26**	.27**	.17*	.24**	.25**	.25**		.53**	.38**	.43**	.40**	.27**	.35**	.16*	.48	.62
a2	.09	.15*	.07	.18*	.11	.16*	.09	.23**	.18*	.09	.43**		.64**	.67**	.55**	.46**	.41**	.28**	.63	.54
a3	.05	.23**	.16*	.28**	.16*	.28**	.19**	.24**	.27**	.20**	.46**	.64**		.77**	.62**	.59**	.35**	.43**	.71	.66
a4	.06	.20**	.15*	.29**	-.02	.25**	.10	.14	.20**	.17*	.33**	.53**	.51**		.69**	.64**	.44**	.46**	.72	.56
a5	.07	.11	.08	.24**	.08	.16*	.15*	.32**	.16*	.08	.42**	.43**	.50**	.50**		.58**	.37**	.44**	.66	.55
a6	.06	.19*	.06	.22**	.05	.22**	.15*	.18*	.22**	.15*	.49**	.48**	.64**	.55**	.55**		.48**	.45**	.60	.62
a7	.11	.10	.14	.18*	.04	.17*	.19*	.12	.19**	.13	.22**	.30**	.26**	.43**	.31**	.45**		.48**	.44	.47
a8	.07	.07	.07	.13	.07	.19*	.12	.16*	.21**	.14	.29**	.29**	.34**	.26**	.38**	.37**	.57**		.48	.48
															Percentage of variance explained				34.11%	29.04%
															Cronbach's Alpha				.88	.85

The British Columbia correlation matrix is above the diagonal, and the Ontario correlation matrix is below the diagonal.
*$p \leq .05$; **$p \leq .01$.

All the items were summed to create the teaching numbers and algebra variable. The descriptive statistics for this variable in British Columbia and Ontario, and the t-test for the difference between means in the two provinces, are reported in Table 16. The distribution of this variable was slightly leptokurtic in both provinces. In addition, the time that teachers spent on teaching numbers and algebra was similar in these two provinces (t=.14, p>.05).

Table 16

Descriptive Statistics for the Teaching Number and Algebra Variable

Province	Range	Mean	SD	Skewness	Kurtosis	Missing	t-test
British Columbia	15–54	36.33	5.59	.09	1.58	5.5%	.14
Ontario	7–54	36.25	6.02	-.28	2.74	8.8%	

Teaching geometry. There were potentially 14 items in the teaching geometry variable as follows:

1. Angles—acute, right, straight, obtuse, reflex.
2. Relationships for angles at a point, angles on a line, vertically opposite angles, angles associated with a transversal cutting parallel lines, and perpendicularity.
3. Properties of geometric shapes: triangles, quadrilaterals, and other common polygons.
4. Construct or draw triangles and rectangles of given dimensions.
5. Congruent figures (triangles, quadrilaterals) and their corresponding measures.
6. Similar triangles and recall their properties.
7. Relationships between two-dimensional and three-dimensional shapes.

8. Pythagorean theorem (not proof) to find length of a side.
9. Measurement, drawing, and estimation of the size of angles, the lengths of lines, areas, and volumes.
10. Measurement formulas for perimeters, circumferences, areas of circles, surface areas, and volumes.
11. Measures of irregular or compound areas (e.g., by covering with grids or dissecting and rearranging pieces).
12. Cartesian plane—ordered pairs, equations, intercepts, intersections, and gradient.
13. Line and rotational symmetry for two-dimensional shapes.
14. Translation, reflection, and rotation.

The correlation coefficients, factor loadings, percentages of variance explained, and reliability coefficients of these items in British Columbia and Ontario are reported in Table 17. Though the correlations varied, all the items loaded highly on a single factor suggesting that they measured the same construct.

All fourteen items were summed to create the teaching geometry variable. The descriptive statistics for this variable in British Columbia and Ontario, and the t-test for the difference between means in the two provinces, are reported in Table 18. The distribution of this variable was almost normal in both provinces. Nevertheless, there was a significant difference between these two provinces suggesting that teachers in Ontario taught more geometry content than teachers in British Columbia ($t=12.98$, $p \leq .001$).

Table 17

Correlation and Factor Analyses of Items for the Teaching Geometry Variable

Correlations															Factor loadings	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	BC	ON
1		.68**	.58**	.48**	.42**	.35**	.31**	.28**	.43**	.32**	.36**	.09	.32**	.38**	.69	.67
2	.44**		.57**	.48**	.47**	.46**	.40**	.35**	.48**	.26**	.20*	.10	.37**	.43**	.74	.41
3	.54**	.32**		.60**	.53**	.46**	.44**	.36**	.38**	.41**	.35**	.11	.38**	.29**	.76	.72
4	.38**	.16*	.45**		.53**	.41**	.45**	.26**	.44**	.36**	.29**	.03	.24**	.17*	.68	.66
5	.33**	.27**	.46**	.59**		.69**	.39**	.22**	.25**	.29**	.26**	.23**	.29**	.29**	.67	.71
6	.38**	.16*	.49**	.37**	.66**		.49**	.24**	.29**	.38**	.26**	.26**	.33**	.31**	.71	.72
7	.17*	.07	.44**	.31**	.35**	.44**		.35**	.50**	.48**	.39**	.31**	.31**	.27**	.69	.57
8	.24**	.19**	.23**	.19**	.20**	.30**	.15*		.43**	.41**	.23**	.28**	.33**	.22**	.53	.44
9	.36**	.26**	.30**	.37**	.25**	.30**	.28**	.22**		.46**	.42**	.25**	.20**	.22**	.64	.58
10	.14	.12	.19*	.26**	.11	.14	.10	.10	.39**		.33**	.15	.28**	.29**	.61	.36
11	.24**	.07	.12	.20**	.19**	.22**	.24**	.18*	.22**	.15*		.31**	.22**	.25**	.51	.42
12	.18*	.10	.01	-.06	.11	.07	.03	.13	-.04	.08	.22**		.35**	.23**	.36	.23
13	.17*	-.08	.14	.15*	.22**	.28**	.26**	.13	.12	.09	.31**	.34**		.73**	.59	.44
14	.18*	-.03	.28**	.17*	.18*	.28**	.24**	.18*	.11	.10	.17*	.23**	.59**		.57	.45
Percentage of variance explained															40.17%	29.98%
Cronbach's Alpha															.88	.80

The British Columbia correlation matrix is above the diagonal, and the Ontario correlation matrix is below the diagonal.
**p* ≤ .05; ***p* ≤ .01.

Table 18

Descriptive Statistics for the Teaching Geometry Variable

Province	Range	Mean	SD	Skewness	Kurtosis	Missing	t-test
British Columbia	14–38	21.09	5.77	.53	-.46	7.1%	12.98***
Ontario	14–42	28.49	5.02	-.16	-.19	7.8%	

***$p \le .001$.

Teaching data and chance. Seven items potentially measured the teaching data and chance variable:

1. Reading data from tables, pictographs, bar graphs, pie charts, and line graphs.
2. Organizing and displaying data using tables, pictographs, bar graphs, pie charts, and line graphs.
3. Characteristics of data sets including mean, median, range, and shape of distribution (in general terms).
4. Interpreting data sets (e.g., draw conclusions, make predictions, and estimate values between and beyond given data points).
5. Data displays that could lead to misinterpretation (e.g., inappropriate grouping and misleading or distorted scales).
6. Using data from experiments to predict chances of future outcomes.
7. Using the chances of a particular outcome to solve problems.

The correlation coefficients, factor loadings, percentages of variance explained, and reliability coefficients of these items in British Columbia and Ontario are reported in Table 19. Though the correlation coefficients varied, especially in British Columbia, all items were consistently and highly loaded on a single factor suggesting that they measured the same construct, which explained more variance for teachers in British Columbia than in Ontario.

Table 19

Correlation and Factor Analyses of Items for the Teaching Data and Chance Variable

Correlations								Factor loadings	
Items	1	2	3	4	5	6	7	BC	ON
1		.85**	.62**	.52**	.43**	.45**	.37**	.73	.75
2	.76**		.66**	.64**	.46**	.46**	.46**	.78	.74
3	.45**	.52**		.71**	.65**	.57**	.56**	.83	.69
4	.46**	.44**	.38**		.67**	.66**	.66**	.85	.72
5	.31**	.33**	.40**	.59**		.77**	.72**	.83	.66
6	.30**	.27**	.31**	.31**	.29**		.82**	.83	.66
7	.28**	.24**	.28**	.34**	.33**	.82**		.80	.66
							Percentage of variance explained	65.44%	48.66%
							Cronbach's Alpha	.91	.82

The British Columbia correlation matrix is above the diagonal, and the Ontario correlation matrix is below the diagonal.
$^{**}p \leq .01$.

All seven of the items were summed to create the teaching data and chance variable. The descriptive statistics for this variable in British Columbia and Ontario, and the t-test for the difference between means in the two provinces, are reported in Table 20. The distribution of this variable was slightly skewed to the right. Additionally, like the teaching geometry variable, teachers in Ontario taught significantly more data and chance content than teachers in British Columbia ($t=10.21$, $p \leq .001$).

Table 20

Descriptive Statistics for the Teaching Data and Chance Variable

Province	Range	Mean	SD	Skewness	Kurtosis	Missing	t-test
British Columbia	7–21	10.28	3.66	1.03	.53	7.1%	10.21***
Ontario	7–21	13.91	3.04	.53	.30	8.3%	

***$p \leq .001$.

Teacher's assignment of homework. Homework are tasks that are assigned to students, and they are meant to be completed during nonschool hours, preferably at home. The majority of teachers, students, and parents are convinced that homework is valuable (Cooper, 2007; Xu, 2005). Reviews of the research on homework (Cooper, 1989; Cooper, Robinson, & Patall, 2006; Cooper & Valentine, 2001; Hattie & Clinton, 2001; Keith, 1986; Walberg, 1991) generally support the view that a moderate amount of time spent by students on homework is associated with achievement gains. The value of homework lies in the distributed practice that it provides for skills that are taught in school. Cooper and his colleague (2006) note that the evidence supporting the relationship between the homework that students complete and their academic achievement is rather weak. One reason is that while students' effort in doing homework can be assessed, the quality or the frequency of that homework generally has not been assessed (see Cooper et al., 2006; Cooper, Lindsay, Nye, & Greathouse, 1998; Muhlenbruck, Cooper, Nye, & Lindsay, 2000; Trautwein & Köller, 2003). A few studies, however, show that a greater number of homework tasks (De Jong, Westerhof, & Creemers, 2000) and more frequent homework tasks (Trautwein, Köller, Schmitz, & Baumert, 2002; Trautwein, 2007) are associated with students' achievement gains. Additionally, compared to the students' evaluation of their homework, teachers normally have a more objective judgment about the amount and the quality of the assignments they give to their students.

As such, in this dissertation the time students spent on homework was evaluated from the students' perspective, and the frequency and quality of assignments were evaluated from the teachers' perspective. For the teachers, the assignment of homework variable was constructed from three items in the teacher questionnaire: "How often do you assign the following kinds of mathematics homework to the TIMSS class?" The responses were made on a 3-point scale.

Always or almost always	Some-times	Never or almost never
1	2	3

1. Doing problem/question sets.
2. Gathering data and reporting.
3. Finding one or more applications of the content covered.

The responses to these items were reverse scored so that higher values indicated more frequent and more challenging homework assigned to students. The correlation coefficients between these items are reported in Table 21. Because the first item "Doing problem/question sets" was slightly negatively related to the other two items, it was dropped from the scale.

Consequently, the teacher's assignment of homework variable was the sum of the second and third items. The descriptive statistics for this variable in British Columbia and Ontario, and the t-test for the difference between means in the two provinces, are reported in Table 22. It is noticeable that teachers in Ontario assigned more homework to their students than teachers in British Columbia (t=6.35, $p \leq .001$). The percentage of missing values in this variable was comparably high in both provinces.

Table 21

Correlation Analysis of Items for the Teacher's Assignment of Homework Variable

Items	1	2	3
1		-.08	-.19*
2	-.05		.38**
3	.01	.30**	

The British Columbia correlation matrix is above the diagonal, and the Ontario correlation matrix is below the diagonal.
*$p \leq .05$; **$p \leq .01$.

Table 22

Descriptive Statistics for the Teacher's Assignment of Homework Variable

Province	Range	Mean	SD	Skewness	Kurtosis	Missing	t-test
British Columbia	2–6	3.14	.96	.31	-.76	11.0%	6.35***
Ontario	2–6	3.78	.91	.05	.23	10.7%	

***$p \leq .001$.

Teacher's attitudes towards teaching. Porter and Freeman (1986) defined teachers' attitudes towards teaching as their specific attitudes towards students, the learning process, the curriculum and pedagogy, the role of schools, and the things they do as teachers. Thus, these predispositions may include questions about the purpose of schooling, about teacher responsibilities for helping their students achieve academic success, and their beliefs that students are capable of achieving these goals. Theoretically, the teachers' attitudes towards teaching are linked closely to their classroom behaviours and practices. Teachers make constant decisions in their classrooms, and their beliefs, attitudes, and priorities provide the framework for these important decisions (Calderhead, 1996; Ernest, 1989; Jackson, 1968; Lortie, 1975; Nespor, 1987; Richardson, 1994).

Though it seems logical to believe that the different attitudes teachers hold result in different outcomes, the positive association between the attitudes of teachers and the students' achievement is not consistently supported by the research literature. For example, the correlation between teachers' satisfaction and their teaching performance is relatively low (Brayfield & Crockett, 1955; Caprara, Barbaranelli, Steca, & Malone, 2006; Laffaldano & Muchinsky, 1985). Nevertheless, teachers' expectations for their students are theoretically important. Teachers' expectations are the inferences that they make about the future behavior and academic success of their students (Raudenbush, 1984; Rosenthal, 1991). The literature suggests that effective teachers not only hold higher expectations, but also they act on their expectations by setting higher goals for their students. Such teachers respond to mistakes during class with appropriate feedback and instruction rather than by lowering their academic standards (Brookover, Beady, Flood, Schweitzer, & Wisenbaker, 1979; Edmonds, 1979; Jussim & Harber, 2005; Rutter, Maughan, Mortimore, Ouston, & Smith, 1979). Nevertheless, there has been considerable debate about the influence that teachers' expectations have on their students' academic, social, and emotional behavior (Babad, 1993; Brophy, 1998; Dusek, 1985; Jussim, Smith, Madon, & Palumbo, 1998).

In this dissertation, the teacher's attitudes variable was constructed from four items the mathematics teachers answered to the question: "How would you characterize each of the following within your school?" The responses were made on a 5-point scale.

Very high	High	Medium	Low	Very low
1	2	3	4	5

1. Teachers' job satisfaction.
2. Teachers' understanding of the school's curricular goals.
3. Teachers' degree of success in implementing the school's curriculum.
4. Teachers' expectations for student achievement.

The responses to these items were reverse coded so that higher values indicated more positive attitudes that teachers had. The correlation coefficients, factor loadings, percentages of variance explained, and reliability coefficients of these items in British Columbia and Ontario, are reported in Table 23. The relatively high correlation coefficients suggested that the items measured the same factor, which is confirmed by the high and consistent factor loadings of the items on a single factor in both British Columbia and Ontario. Additionally, the scales were reliable as indicated by the high Cronbach's Alpha reliability coefficients in both provinces.

These items were summed to create the teacher's attitudes towards teaching variable. Table 24 reports the descriptive statistics for this variable in British Columbia and Ontario, and the t-test for the difference between means in the two provinces. In both provinces, mathematics teachers generally had positive attitudes, and there was no difference between the average attitudes in the two provinces as indicated by the t-test. The distribution of this variable was normal in both provinces.

Table 23

Correlation and Factor Analyses of Items for the Teacher's Attitudes towards Teaching Variable

Correlations						Factor loadings	
Items	1	2	3	4		BC	ON
1		.30**	.35**	.34**		.61	.72
2	.48**		.59**	.49**		.80	.82
3	.47**	.60**		.53**		.83	.85
4	.37**	.50**	.59**			.79	.78
					Percentage of variance explained	58.12%	62.76%
					Cronbach's Alpha	.75	.80

The British Columbia correlation matrix is above the diagonal, and the Ontario correlation matrix is below the diagonal.
$**p \leq .01$.

Table 24

Descriptive Statistics for the Teacher's Attitudes towards Teaching Variable

Province	Range	Mean	SD	Skewness	Kurtosis	Missing	t-test
British Columbia	10–20	15.60	1.97	-.07	.00	7.1%	.95
Ontario	10–20	15.80	2.17	-.04	-.19	3.4%	

Homogeneity of students in classroom. Today, teachers must deal with the increased heterogeneity of students in their classrooms. The diversity of student body is increasing because of the inclusion of more at-risk and disabled students. Though some excellent teachers can be effective with heterogeneous classrooms (Wright, Horn, & Sanders, 1997; Sanders & Horn, 1998; Sanders & Rivers, 1996), perhaps the majority of teachers can only handle a narrow band of heterogeneity in their students (Kulik, 1992; Kulik & Kulik, 1982). Additionally, having less heterogeneity of students is more important in subjects in which the skills develop in a hierarchical fashion (e.g., mathematics) than in subjects in which the skills are less clearly dependent on the mastery of earlier material (e.g., social studies). In these hierarchically organized subjects, if the teacher proceeds too rapidly, some students will fail to develop the prerequisite skills, whereas if teachers take the time needed to ensure that all students have the prerequisite skills, the more able students will become bored and waste a great deal of time. For this reason, the greater homogeneity of students in classroom probably has a positive impact on the students' achievement in mathematics.

In this dissertation, the homogeneity of students that the mathematics teachers had in their classrooms was assessed on five items in response to: "In your view, to what extent do the following limit how you teach the TIMSS class?" The responses were made on a 5-point scale.

Not applicable	Not at all	A little	Some	A lot
1	2	3	4	5

1. Students with different academic abilities.
2. Students who come from a wide range of backgrounds (e.g., economic, language).
3. Students with special needs (e.g., hearing, vision, speech impairment, physical disabilities, mental or emotional/psychological impairment).
4. Uninterested students.
5. Disruptive students.

The responses to these items were reverse scored so that higher values indicated higher student homogeneity in the teachers' classrooms. The correlation coefficients, factor loadings, percentages of variance explained, and reliability coefficients of these items in British Columbia and Ontario, are reported in Table 25. All correlation coefficients were relatively strong except for the correlation between the second and fifth items in the Ontario sample (r=.08). Nevertheless, all the items were consistently and highly loaded on a single factor in both British Columbia and Ontario, and they had adequate reliabilities.

Table 25

Correlation and Factor Analyses of Items for the Homogeneity of Students in Classroom Variable

Correlations						Factor loadings	
Items	1	2	3	4	5	BC	ON
1		.44**	.32**	.46**	.38**	.73	.80
2	.33**		.37**	.31**	.22**	.64	.50
3	.45**	.29**		.34**	.29**	.62	.60
4	.54**	.29**	.30**		.70**	.82	.85
5	.39**	.08	.16*	.68**		.75	.70
					Percentage of variance explained	51.21%	48.90%
					Cronbach's Alpha	.75	.72

The British Columbia correlation matrix is above the diagonal, and the Ontario correlation matrix is below the diagonal.
**$p \leq .01$.

These items were summed to create the homogeneity of students in classroom variable. Table 26 reports the descriptive statistics for this variable in British Columbia and Ontario, and the t-test for the difference between means in the two provinces. This variable was normally distributed in both provinces, and students in Ontario seemed to be slightly more homogeneous than students in British Columbia. The difference, however, was not significant (t=1.83, p>.05).

Table 26

Descriptive Statistics for the Homogeneity of Students in Classroom Variable

Province	Range	Mean	SD	Skewness	Kurtosis	Missing values	t-test
British Columbia	5–24	13.54	3.36	.10	.17	8.2%	1.83
Ontario	5–25	14.22	3.52	.00	.18	11.2%	

Physical resources in classroom. As noted previously, if principals can provide a meaningful evaluation of the physical resources that teachers need for instructional purposes, teachers themselves can provide an even more meaningful assessment about these resources. In this dissertation, the physical resources that are available to teachers in their classrooms were derived from the mathematics teachers' responses to three items: "In your current school, how severe is each problem?" The responses were made on a 3-point scale.

Not a problem	Minor problem	Serious problem
1	2	3

1. The school building needs significant repair.

2. Classrooms are overcrowded.

3. Teachers do not have adequate workspace outside their classrooms.

The responses to these items were reverse coded so that higher values indicated that more physical resources were available. The correlation coefficients, factor loadings, percentages of variance explained, and reliability coefficients of these items in British Columbia and Ontario are reported in Table 27. The items consistently loaded on one factor, between .67 and .80 in British Columbia and between .73 and .82 in Ontario, indicating that they all measured the same construct. The Cronbach's Alpha coefficients for these items were acceptable but relatively low, 0.62 in British Columbia and 0.64 in Ontario.

Table 27

Correlation and Factor Analyses of Items for the Physical Resources in Classroom Variable

Correlations					Factor loadings	
Items	1	2	3		BC	ON
1		.29**	.31**		.67	.74
2	.29**		.46**		.79	.73
3	.4**	.41**			.80	.82
				Percentage of variance explained	56.90%	58.40%
				Cronbach's Alpha	.62	.64

The British Columbia correlation matrix is above the diagonal, and the Ontario correlation matrix is below the diagonal.
**$p \leq .01$.

All items were summed to create the physical resources in classroom variable. The descriptive statistics for this variable in British Columbia and Ontario, and the t-test for the difference between means in the two provinces are reported in Table 28. This variable was almost normally distributed in both British Columbia and Ontario.

Table 28

Descriptive Statistics for the Physical Resources in Classroom Variable

Province	Range	Mean	SD	Skewness	Kurtosis	Missing	t-test
British Columbia	3–9	7.14	1.53	-.48	-.60	6.0%	.76
Ontario	3–9	7.02	1.63	-.67	-.25	3.9%	

Instructional resources in classroom. Similar to the physical resources, the instructional resources available to teachers in their classrooms were assessed based on three items answered by the teachers: "In your view, to what extent do the following limit how you teach the TIMSS class?" The responses were made on a 5-point scale.

Not applicable	Not at all	A little	Some	A lot
1	2	3	4	5

1. Shortage of textbooks for student use.
2. Shortage of other instructional equipment for students use.
3. Shortage of equipment for your use in demonstrations and other exercises.

The responses to these items were reverse scored so that higher values indicated more materials available for teaching. The correlation coefficients, factor loadings, percentages of variance explained, and reliability coefficients of these items in British

Columbia and Ontario, are reported in Table 29. As indicated by the consistent and high factor loadings on one factor, all the items were measuring the same construct. Additionally, these items were reliable as confirmed by the Cronbach's Alpha coefficients.

Table 29

Correlation and Factor Analyses of Items for the Instructional Resources in Classroom Variable

Correlations				Factor loadings	
Items	1	2	3	BC	ON
1		.42**	.29**	.61	.80
2	.64**		.80**	.93	.94
3	.56**	.85**		.89	.91
			Percentage of variance explained	67.34%	78.04%
			Cronbach's Alpha	.75	.85

The British Columbia correlation matrix is above the diagonal, and the Ontario correlation matrix is below the diagonal.
**$p \leq .01$.

All items were summed to create the instructional resources in classroom variable. Table 30 reports the descriptive statistics for this variable in British Columbia and Ontario, and the t-test for the difference between means in the two provinces. This variable was almost normally distributed in both provinces, and slightly platykuritic in British Columbia. The teachers in British Columbia had significantly more instructional resources than the teachers in Ontario (t=3.58, $p \leq .001$).

Table 30

Descriptive Statistics for the Instructional Resources in Classroom Variable

Province	Range	Mean	SD	Skewness	Kurtosis	Missing	t-test
British Columbia	3–15	10.75	2.19	-.78	1.50	7.2%	3.43***
Ontario	3–15	9.85	2.58	-.50	.14	8.6%	

***$p \leq .001$.

Student Variables

Because students respond differentially to the learning opportunities that are provided by their teachers and schools, five student variables, the student's gender, SES (measured by two variables, mother's education and number of books in home), instrumental motivation, educational expectations, and effort in completing homework, were used in this dissertation. All variables were obtained from responses in the questionnaire answered by students.

Gender. Gender differences in academic achievement have been extensively studied, and it is widely accepted that, on average, females score higher than males on verbal tests (Halpern, 2000; Hyde & Linn, 1988; Stumpf, 1995), and males score higher than females on mathematical tests (Halpern, 2000; Hyde, Fennema, & Lamon, 1990; Stumpf, 1995) and spatial tests (Beller & Gafni, 1996; Hedges & Nowell, 1995; Voyer, Voyer, & Bryden, 1995). The gender difference is so evident that people have developed different expectations about male and female students' performances in mathematics (Clifton, Perry, Parsonson, & Hryniuk, 1986). However, more recently, gender differences in mathematics have been relatively small (Tate, 1997). For example, Friedman (1989) reported, on average, a regression coefficient of .024 in favor of males.

In this dissertation, male students were coded “0” and females were coded “1”. In British Columbia, 52.5% of the students were female and in Ontario, 51.6% were female. A chi-square test indicated that there were no differences between the two provinces in the percentage of female students.

Socioeconomic status (SES). Socioeconomic status (SES) is probably the most widely used variable in educational research, either as a predictor of performances, as a control variable, or as a way of stratifying samples (White, 1982). In general, SES describes an individual’s and family’s ranking on a hierarchy of access to wealth and educational resources (Mueller & Parcel, 1981). Normally, the family SES helps to determine the environment to which students are exposed and the expectations families have for their children’s education (Reynolds & Walberg, 1992). Also, considerable research shows that the family SES indirectly affects the quality of the instruction that students often receive in schools. Not surprisingly, of all the student background variables examined in the literature, family SES has one of the strongest correlates with students’ academic achievement. The correlation between the SES and the academic achievement, around .30, has been consistently found among a large number of published studies (see literature reviews by Sirin, 2005 and White, 1982).

In this dissertation, two variables, the mother’s education and the number of books in the home, were used to indicate the SES of the students. The parental occupation and income, two indicators of the student SES, were not used because they were not measured in the Canadian TIMSS 2007.

Mother's education. The amount of education attained by the students' mothers was derived from the students' responses to an item asking about their mothers' education, which was coded on the International Standard Classification of Education index as follows (the explanations by TIMSS are in parentheses):

1 ISCED 1 or did not go to school (primary education, first stage of basic education and below)

2 ISCED 2 (lower secondary or second stage of basic education)

3 ISCED 3 (upper secondary education)

4 ISCED 4 (post-secondary non-tertiary education)

5 ISCED 5B (tertiary education, shorter than those in 5A and focus on occupationally specific skills geared for entry into the labour market)

6 ISCED 5A, first degree (tertiary programmes intended to provide sufficient qualifications for gaining entry into advanced research programmes and profession with high skills requirements)

7 Beyond ISCED 5A, first degree.

Table 31 reports the descriptive statistics for the mother's education in British Columbia and Ontario, and the t-test for the difference between means in the two provinces. While the average education that the students' mothers attained was high in both provinces, only slightly below the university level, mothers in Ontario had, on average, significantly more education than mothers in British Columbia ($t=2.99$, $p \leq .01$).

Table 31

Descriptive Statistics for the Mother's Education Variable

Province	Range	Mean	SD	Skewness	Kurtosis	Missing	t-test
British Columbia	1–7	4.72	1.76	.05	-1.47	0	2.99**
Ontario	1–7	4.87	1.69	-.01	-1.41	0	

**$p \leq .01$.

Number of books in home. The number of books in the students' homes was derived from asking them about the books in their homes, excluding magazines, newspapers, and school books. The students were given five options as follows:

1 None or very few.

2 Enough to fill one shelf (11–25 books).

3 Enough to fill one bookcase (26–100 books).

4 Enough to fill two bookcases (101–200 books).

5 Enough to fill three or more bookcases (more than 200 books).

The descriptive statistics for this variable in British Columbia and Ontario, and the t-test for the difference between means in the two provinces, are reported in Table 32. The distribution of this variable was almost normal in both provinces. Additionally, students in British Columbia reported significantly more books in their homes than students in Ontario (t=3.84, $p \leq .001$).

Table 32

Descriptive Statistics for the Number of Books in Home Variable

	Range	Mean	SD	Skewness	Kurtosis	Missing	t-test
British Columbia	1–5	3.42	1.23	-.27	-.88	.2%	3.84***
Ontario	1–5	3.28	1.22	-.12	-.89	.2%	

***$p \leq .001$.

Instrumental motivation. Instrumental motivation indicates the extent to which school tasks are perceived as being important for the students in attaining their goals. There is a convincing body of research validating the notion of instrumental motivation and claiming that tasks that are important for the attainment of future goals provide greater incentive to people; thus, engaging in these instrumental tasks is more vigorous and the achievement is more assured (see Gardner & MacIntyre, 1991; Miller, Greene, Montalvo, Ravindran, & Nichols, 1996).

In this dissertation, the instrumental motivation variable was created from four items in the students' questionnaire in response to: "How much do you agree with these statements about mathematics?" The responses were made on a 4-point scale.

Agree a lot	Agree a little	Disagree a little	Disagree a lot
1	2	3	4

1. I think learning mathematics will help me in my daily life.
2. I need mathematics to learn other school subjects.
3. I need to do well in mathematics to get into the university or college of my choice.
4. I need to do well in mathematics to get the job I want.

The responses to these four items were reverse scored so that higher values indicated higher levels of motivation. The correlation coefficients, factor loadings, percentages of variance explained, and reliability coefficients of these items in British Columbia and Ontario, are reported in Table 33. The comparably high correlation coefficients between the items and the consistent loadings on one factor showed that they were measuring the same construct. Additionally, these items were adequately reliable as indicated by the Cronbach Alpha coefficients.

Table 33

Correlation and Factor Analyses of Items for the Instrumental Motivation Variable

	Correlations					Factor loadings	
Items	1	2	3	4		BC	ON
1		.51**	.36**	.39**		.75	.71
2	.42**		.35**	.35**		.73	.70
3	.31**	.33**		.51**		.74	.75
4	.38**	.32**	.52**			.75	.77
					Percentage of variance explained	55.46%	53.46%
					Cronbach's Alpha	.73	.71

The British Columbia correlation matrix is above the diagonal, and the Ontario correlation matrix is below the diagonal.
$^{**}p \leq .01$.

These items were summed to create the students' instrumental motivation variable. The descriptive statistics for this variable in British Columbia and Ontario and the t-test for the difference between means in the two provinces are reported in Table 34. The distribution of this variable was slightly skewed to the left and leptokurtic in both British

Columbia and Ontario. In addition, students in Ontario had, on average, significantly higher levels of instrumental motivation than students in British Columbia (t=4.68, $p \leq .001$).

Table 34

Descriptive Statistics for the Instrumental Motivation Variable

	Range	Mean	SD	Skewness	Kurtosis	Missing	t-test
British Columbia	4–16	13.40	2.29	-1.07	1.22	2.3%	4.68***
Ontario	4–16	13.71	2.09	-.99	.98	1.1%	

***$p \leq .001$.

Educational expectations. Expected educational attainment reflects the students' subjective assessment of the probability that they will obtain a certain level of education in the future. Expectations are future-oriented because successful performances on current educational tasks do not, in themselves, produce the desired levels of academic attainment. Rather, the future goals students hold provide the motivation to strive for their goals. When people initially commit themselves to future goals, they are thought to purposefully generate a coherent framework of proximal sub-goals to guide them along the way (Ames, 1992; Butler, 1993; Dweck & Leggett, 1988; Pintrich, 2000; Utman, 1997). Not surprisingly, after controlling for past performances, educational expectations have been shown to predict real-world performances on examinations and course grades (Dweck & Sorich, 1999; Kaplan & Maehr, 1999; Marjoribanks, 2005; Roeser, Midgely, & Urdan 1996).

In this dissertation, the students' educational expectations were collected from their responses to a single item asking about their educational expectations, which was classified on the International Standard Classification of Education index and coded as follows (the explanations by TIMSS are in parentheses):

1 Finish ISCED 3 (upper secondary education)

2 Finish ISCED 4 (post-secondary non-tertiary education)

3 Finish ISCED 5B (tertiary education, shorter than those in 5A and focus on occupationally specific skills geared for entry into the labour market)

4 Finish ISCED 5A, first degree (tertiary programmes intended to provide sufficient qualifications for gaining entry into advanced research programmes and profession with high skills requirements)

5 Beyond ISCED 5A

The descriptive statistics for this variable in British Columbia and Ontario, and the t-test for the difference between means in the two provinces, are reported in Table 35. The distribution of this variable was slightly skewed to the left in both provinces. Students in Ontario, however, had significantly higher educational expectations than students in British Columbia (t=4.29, $p \leq .001$).

Table 35

Descriptive Statistics for the Educational Expectations Variable

Province	Range	Mean	SD	Skewness	Kurtosis	Missing	t-test
British Columbia	0–5	3.67	1.67	-1.02	-.36	1.3%	4.29***
Ontario	0–5	3.88	1.55	-1.28	.39	.5%	

***$p \leq .001$.

Effort in completing homework. The students' effort in doing their homework was evaluated by the number of minutes they spent doing homework in mathematics when homework was assigned. The data was collected from the students' responses to the question: "When your teacher gives you mathematics homework, about how many minutes do you usually spend on your homework?" The responses were coded as follows:

1 1–15 minutes

2 16–30 minutes

3 31–60 minutes

4 61–90 minutes

5 More than 90 minutes

The descriptive statistics for this variable in British Columbia and Ontario, and the t-test for the difference between means in the two provinces, are reported in Table 36. The distribution of this variable was almost normal, and students from British Columbia spent significantly more time doing homework than students from Ontario (t=6.60, $p \leq .001$).

Table 36

Descriptive Statistics for the Effort in Completing Homework Variable

Province	Range	Mean	SD	Skewness	Kurtosis	Missing	t-test
British Columbia	1–5	2.32	.881	.437	.215	5.1%	6.60***
Ontario	1–5	2.14	.860	.604	.452	4.1%	

***$p \leq .001$.

The sample and the variables used in this dissertation have now been examined. Along with the description of those variables, the research literature supporting these variables for the analyses that are presented in Chapter 4 has been reviewed. In the next section, four hypotheses that were tested in this study are presented. They were derived from the research literature and the descriptive statistics.

Hypotheses

Hypothesis 1. The aggregate analysis gives the most biased estimates of school effectiveness on the students' academic achievement.

Hypothesis 2. The HLM gives the most accurate estimates of school effectiveness on the students' academic achievement.

Hypothesis 3. There are significant provincial differences between British Columbia and Ontario in the students' academic achievement.

Hypothesis 4. The physical resources of schools, which are evaluated by school principals and classroom teachers, have a positive impact on the students' academic achievement in both British Columbia and Ontario.

Summary

The sample of schools, teachers, students, and the variables included in this dissertation have been described in this chapter. More specifically, the first section describes the Canadian TIMSS 2007 data gathered in British Columbia and Ontario. In British Columbia, 134 schools, 169 teachers, and 2,392 students were sampled. In Ontario, 169 schools, 198 teachers, and 2,080 students were sampled. In the second section, the dependent variable, the students' academic achievement in mathematics, is described in detail. Achievement was evaluated by the five plausible values for the students' mathematics scores in the TIMSS data. In the third section, four school variables (the school's physical resources, instructional resources, average SES of students, and administration), eight teacher and classroom variables (the teacher's gender, experience, instructional time measured by teaching numbers and algebra, geometry, and data and chance, assignment of homework, attitudes towards teaching, homogeneity of students, and physical and instructional resources in classroom), and five student variables (the student's gender, SES assessed by mother's education and number of books in home, instrumental motivation, educational expectations, and effort in completing homework) are described. All these variables adequately meet the required assumptions for the analyses that are presented in the next chapter. As noted, there are some important differences in the students' mathematics achievement between British Columbia and Ontario. Such differences, of course, suggest that the multivariate analysis presented in Chapter 4 may illustrate important differences between the two provinces. The review of literature and the descriptive statistics are the basis of the four hypotheses that were tested in this study. Chapter 4 reports how the effects of the independent variables on the dependent variable were estimated using different statistical procedures for the hierarchical data: students nested in schools, and schools nested in provinces.

CHAPTER 4

RESULTS

This chapter has four sections. The first section reports an examination of missing values in some of the variables and the collinearity among all the independent variables. The second section reports results of the aggregate, disaggregate, and HLM procedures. The effects of the school, teacher/classroom, and student variables on the students' academic achievement in mathematics in these procedures are reported and compared. The comparisons of these three procedures clearly show that HLM has more appropriate standard errors and more accurate coefficients. More specifically, a two-level HLM analysis is much better than the aggregate analysis and slightly better than the disaggregate analysis in evaluating the effectiveness of schools in British Columbia and Ontario. The third section evaluates the HLM analysis in greater detail. Finally, the results are interpreted in terms of the hypotheses that were specified previously in Chapter 3.

Testing Two Assumptions

Parametric statistics have basic assumptions about data, and it is important to ensure the data do not violate these assumptions. The distribution of each variable is examined in Chapter 3 and they are all roughly normally distributed. This section reports the examination of missing values to determine the degree to which the missing values could have affected the results of analysis. Following this, I examine the collinearity among the independent variables to ensure that this condition is not a problem.

Missing Values

Missing values are common in social sciences (Little & Rubin, 1989), and it is no exception in the TIMSS data. In fact, there are missing values in the data for schools, teachers/classrooms, and students. One important limitation of the TIMSS data is that a few schools do not have any school and teacher information. In other words, the principals' answers to the questions about their schools and teachers are missing. After excluding these few schools in this analysis, the missing values for other teacher/classroom variables were below 10%. Nevertheless, there was still a significant amount of missing values in the SES information that was obtained from the students. Accordingly, the effects of missing values in the school and student data were assessed.

Among the samples reported in Table 4, 16 schools (10.67% out of 150 schools) in British Columbia with 396 students (9.30% of the students) and seven schools (3.98% out of 176 schools) in Ontario with 113 students (3.28% of the students) were missing school-level values. If these schools were random, the effect of the missing values would not be a problem in the analysis. I assessed the effect of missing values on the analysis by comparing the students in the schools with missing school-level values to the students in the schools with no missing values.

Table 37

Comparisons between Students from Schools With and Without Missing Values in British Columbia and Ontario

	Not missing		Missing		Mean Difference	T-test
A. British Columbia	Mean	SD	Mean	SD		
MathPV1	508.10	70.84	492.82	69.89	15.28	4.09***
MathPV2	507.99	71.84	490.63	72.44	17.36	4.58***
MathPV3	508.39	72.20	492.88	71.67	15.50	4.07***
MathPV4	508.35	71.17	493.62	68.70	14.72	3.93***
MathPV5	508.85	70.81	491.33	70.71	17.52	4.69***
Gender	.51	.50	.46	.50	.05	1.69
Mother education	4.72	1.76	4.61	1.66	.10	.86
Father education	4.78	1.83	4.69	1.84	.09	.74
Edu. expectations	3.32	1.85	3.30	1.75	.02	.21
No. of Books	3.36	1.25	3.30	1.22	.04	2.42*
B. Ontario						
MathPV1	515.65	66.63	488.17	87.94	27.48	4.26***
MathPV2	516.44	67.22	490.26	84.87	26.18	4.03***
MathPV3	515.51	67.46	490.02	89.89	25.49	3.90***
MathPV4	515.59	67.81	488.92	90.96	26.68	4.06***
MathPV5	516.06	67.50	487.31	85.55	28.76	4.41***
Gender	.50	.50	.48	.50	.03	.58
Mother education	4.87	1.69	4.46	1.67	.41	2.03*
Father education	4.82	1.77	4.51	1.75	.31	1.34
Edu. expectations	3.58	1.74	2.92	1.82	.66	3.79***
No. of Books	3.20	1.22	2.82	1.23	.38	3.15**

MathPV, plausible values of students in mathematics.
**p* < .05, ***p* < .01, ****p* < .001.

First, the student samples were divided into two groups: students from schools with school-level data and students from schools with missing school-level data. Second, the students' mean scores in mathematics, represented by the five plausible values, and the mean scores on five other student-level variables were compared. The differences between the two groups were assessed using t-tests. The results of these comparisons in British Columbia and Ontario are reported in Table 37.

In both provinces, the students coming from schools without missing values had significantly higher achievement scores than the students from schools with missing values. Specifically, in British Columbia, students from schools without missing values had, on average, 16.08 points higher scores on the mathematics test than students from schools with missing values. In Ontario, the difference was even larger; specifically, students from schools without missing values had, on average, 26.92 points higher scores than students from schools with missing values. Nevertheless, in British Columbia the two groups of students did not differ in their demographic characters and educational expectations. Though there was a statistically significant difference on their number of books in their homes ($t=2.42$, $p<.05$), the difference between their means was small. In Ontario, on the other hand, there were significant differences between the two groups of students on their mothers' education, their own educational expectations, and the number of books in their homes. Particularly, students from schools without missing values had significantly higher levels of mother's education, higher educational expectations, and more books in their homes than students from schools with missing values.

As a result of these analyses, 16 schools and their 396 students in British Columbia, and seven schools and their 113 students in Ontario were excluded from the analyses in the next section of this chapter. Given the significant differences between schools with and without missing data, the effect of excluding these schools and their students is discussed in Chapter 5.

For the remaining students, five individual variables were examined and both the mother's and the father's education had relatively large amounts of missing values, which was probably caused by the students' checking "I don't know" to the question about their parents' education. To find out if students who had missing values on their parental education were different from those without missing values, the students were split into two groups depending on if they had missing values on their mothers' education in both British Columbia and Ontario. The percentages of missing values on five student variables in these two provinces are reported in Table 38.

Table 38

Percentages of Missing Values in the Students' Data in British Columbia and Ontario

	British Columbia			Ontario		
	Total	With mother edu.	Missing mother edu.	Total	With mother edu.	Missing mother edu.
Gender	1.9%	0	5.0%	2.1%	0	5.5%
Mother education	38.0%	0	100.0%	37.6%	0	100.0%
Father education	39.4%	13.7%	81.2%	41.4%	16.2%	83.2%
No. of books in home	2.3%	.2%	5.6%	2.3%	.3%	5.8%
Educational expectations	3.7%	1.3%	7.6%	2.1%	0	5.5%

Among students, 38.0% (1,468 students) in British Columbia and 37.6% (1,255 students) in Ontario did not report their mothers' education. Additionally, in both provinces, students who reported that they did not know the amount of education their mothers were more likely to report that they did not know their fathers' education. The relationships between not answering the question about their mothers' education and not answering the questions about their own gender, number of books in their homes, and educational expectations were comparably weak. To further compare the differences between students with and without missing values on their mothers' education, their mathematics achievement scores, represented by the five plausible values, were compared. These comparisons are reported in Table 39.

Similar to the previous findings, there were significant differences between the two groups of students. Specifically, in both provinces the students who reported their mothers' education had significantly higher mathematics achievement scores, on average 19.46 and 22.56 points in British Columbia and Ontario respectively, than those who did not report their mothers' education. In addition, students who reported their mothers' education had significantly higher educational expectations than those who did not. Because these findings showed strong missing value effects, the decision to use only students who reported their mothers' education in the analysis is addressed in Chapter 5.

Table 39

Comparisons between Students With and Without Missing Values on Mother's Education in British Columbia and Ontario

	Not missing		Missing		Mean Difference	t-test
A. British Columbia	Mean	SD	Mean	SD		
MathPV1	515.51	69.82	496.04	70.84	19.47	8.36***
MathPV2	515.36	71.35	496.00	71.05	19.35	8.20***
MathPV3	515.65	71.79	496.55	71.33	19.09	8.04***
MathPV4	515.95	70.41	495.95	70.68	20.00	8.56***
MathPV5	516.22	70.19	496.84	70.18	19.39	8.33***
Gender	.52	.50	.48	.50	.04	2.55*
Mother education	4.72	1.76	-	-	-	-
Father education	4.78	1.82	4.80	1.938	-.02	-.13
No. of books	3.42	1.23	3.07	1.26	.35	8.45***
Edu. expectation	3.67	1.67	2.72	1.993	.96	15.69***
B. Ontario						
MathPV1	524.21	66.08	501.48	65.13	22.73	9.68***
MathPV2	524.35	66.60	503.33	66.21	21.03	8.85***
MathPV3	524.36	66.46	500.83	66.56	23.53	9.90***
MathPV4	524.06	66.94	501.55	66.93	22.52	9.41***
MathPV5	524.70	66.50	501.74	66.73	22.97	9.65***
Gender	.52	.50	.48	.50	.03	1.73
Mother education	4.87	1.69	-	-	-	-
Father education	4.82	1.76	4.85	1.81	-.04	-.28
No. of books	3.28	1.22	3.05	1.19	.23	5.06***
Edu. expectation	3.88	1.55	3.05	1.93	.83	13.42***

MathPV, plausible values in mathematics.
*$p \leq .05$, ***$p \leq .001$.

Collinearity

Collinearity, also referred as multicollinearity, happens because two or more independent variables in a regression calculation are highly correlated. When the correlation among independent variables is high (above .70), it is difficult or impossible to determine the exact effect of each independent variable on the dependent variable. Thus, collinearity increases uncertainty in the analysis because the standard errors would be large and the effect parameter would be inaccurate. As such, collinearity makes it difficult, if not impossible, to interpret a particular regression coefficient. Because collinearity happens when two or more independent variables are highly correlated, an examination of the correlation coefficients among the independent variables is recommended as a way of determining if collinearity is a potential problem (Cohen, Cohen, West, & Aiken, 2003).

The correlation matrices for the school and teacher/classroom variables in British Columbia and Ontario are reported in Table 40. It is noted that the largest correlation was between teaching geometry and teaching data and chance in British Columbia (r=.45) and between teaching number and algebra and teaching data and chance in Ontario (r=.48). Physical resources and instructional resources in schools were also highly correlated in both British Columbia (r=.44) and in Ontario (r=.40). Similarly, physical resources and instructional resources in classroom were highly correlated in British Columbia (r=.30) and in Ontario (r=.34). The relationships among the instructional time on number and algebra, geometry, and data and change were strong in both British Columbia (r=.34–.48) and Ontario (r=.30–.45). Finally, among the teacher/classroom variables, the homogeneity of students in classroom, physical resources in classroom, and instructional resources in classroom were comparably highly correlated in British Columbia (r=.16–.34) and Ontario (r=.21–.31). However, none of the correlation coefficients were larger than .50, which suggests that collinearity is not a problem among school and teacher/classroom variables in either British Columbia or Ontario.

Table 40

Correlations among School Variables in British Columbia and Ontario

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Sch. physi. res.		.44**	.01	-.06	.13	-.01	-.14	.05	.07	.06	-.02	.17	.19*	.26**
2. Sch. instruc. res.	.40**		.07	-.04	.08	.15	-.17	.02	.02	-.06	.23**	.21*	.05	.23**
3. Sch. average SES	.09	.13		-.11	-.20*	.04	-.07	-.08	-.02	-.10	.21*	.20*	.06	.07
4. Sch. admin.	.08	.01	.00		.11	.04	-.03	-.13	-.03	-.01	.18*	.03	-.05	.07
5. Tch. gender	-.12	.01	.03	.07		-.13	.08	-.01	-.01	-.04	.03	.25**	-.07	.14
6. Tch. experience	-.09	-.12	-.08	.27**	.00		-.02	.13	-.02	-.04	.13	.01	-.04	.16
7. Tch. number	-.04	-.05	-.06	-.04	-.02	.01		.36**	.30**	.01	.07	.10	.01	.05
8. Tch. geometry	.12	.08	-.01	-.03	-.16*	-.02	.34**		.45**	.07	-.04	.04	.14	-.02
9. Tch. data	.03	.05	-.19*	-.01	-.02	.14	.48**	.43**		.08	.04	.07	.03	.07
10. Tch. homework	-.11	-.05	-.01	-.06	.00	.07	.06	-.07	.06		.09	.10	.03	.04
11. Tch. attitudes	.00	.10	.14	.04	.00	.09	-.01	.07	.12	.23**		.31**	.26**	.21*
12. Cla. homo.	-.13	-.06	.11	.06	-.09	-.01	.08	.11	.09	.07	.26**		.26**	.32**
13. Cla. physi. res.	.14	.06	-.01	.08	.02	.00	.01	-.01	.10	.16*	.33**	.16*		.30**
14. Cla. Instruct. res.	.09	.12	.00	-.01	-.06	-.13	-.07	.06	.03	.18*	.31**	.29**	.34**	

The British Columbia correlation matrix is above the diagonal, and the Ontario matrix is below the diagonal.
*$p \leq .05$, **$p \leq .01$.

There were, however, obvious differences among some variables in the two provinces, which suggested that there might be important differences between British Columbia and Ontario. More specifically, in British Columbia, the physical resources in schools variable had statistically significant and moderately strong correlations with the physical resources in classroom (r=.19) and the instructional resources in classroom (r=.26), while in Ontario the same variable had no significant correlations with these two variables. As well, the instructional resources in school variable had statistically significant and moderately strong correlations in British Columbia with the teacher's attitudes towards teaching variable (r=.23), the homogeneity of students in classroom variable (r=.21), and the instructional resources in classroom variable (r=.23). In Ontario, on the other hand, the instructional resources in school variable had no significant correlations with these three variables. It is interesting that the average SES of students was negatively related to the teacher's gender (r= -.20) in British Columbia, which means that in this province, schools with students who had lower average SES had more female teachers, while the same relationship was virtually zero in Ontario. Finally, the administration of schools had a moderately strong correlation with the teachers' attitudes in British Columbia (r=.18) and a strong correlation with the teachers' years of experience in Ontario (r=.27).

Table 41 presents the correlation coefficients among the student variables in British Columbia and Ontario. The correlation coefficients among these independent variables were either weak or moderately strong, ranging from -.06 to .30. As such, collinearity was not significantly present among student variables in either province.

To assess the correlations between the school-level and the student-level variables together, all school-level variables were distributed to the student level and then the correlations among all the independent variables were examined. The results are reported in Table 42. The correlations among all these variables ranged between -.20 and +.20. Thus, it is concluded that collinearity was not a problem for the full analysis presented later in this chapter.

Table 41

Correlations among Student Variables in British Columbia and Ontario

	1	2	3	4	5	6
1. Gender		-.01	.05*	.00	.08**	.05*
2. Mother education	-.05*		.30**	.10**	.26**	.00
3. Number of books	.01	.29**		.13**	.20**	.06**
4. Instru. motivation	-.06**	.10**	.09**		.28**	.05*
5. Edu. expectations	.10**	.30**	.17**	.22**		.06**
6. Effort in homework	.09**	-.01	.08**	.03	.05*	

The British Columbia correlation matrix is above the diagonal, and the Ontario matrix is below the diagonal.
$^*p \leq .05$, $^{**}p \leq .01$.

In this exploratory data analysis section, two assumptions for both general linear regressions and HLM were assessed. More specifically, the missing values in both the school and the student data were not random. Schools that had missing values on the school questionnaire were also likely to have lower average student achievement scores in both British Columbia and Ontario. Students who did not report their parents' education also had lower achievement scores in both provinces. As a consequence, to minimize the estimated errors, 16 schools and 1,864 students in British Columbia and seven schools and 1,368 students in Ontario were excluded. In addition, preliminary analyses showed that the collinearity was low among all the independent variables in both British Columbia and Ontario. In the next section, the aggregate, disaggregate, and HLM analyses are presented using the remaining data, 134 schools with 2,392 students in British Columbia and 169 schools with 2,080 students in Ontario.

Table 42

Correlations among Independent Variables in British Columbia and Ontario

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1. Sch. physi. res.		.45**	.11**	-.14**	.03	.04	-.17**	.07**	.10**	-.03	-.07**	.13**	.19**	.21**	.04*	.00	.00	.03	.02	-.01
2. Sch. instru. res.	.40**		.07**	-.05*	.07**	.14**	-.20**	.00	.01	-.12**	.19**	.17**	.06**	.13**	.02	.01	.01	.04	.02	.02
3. Sch. SES	.05*	.12**		-.16**	-.23**	.02	-.11**	-.03	.03	-.06**	.18**	.17**	.10**	.00	.04*	.16**	.14**	-.02	.00	-.01
4. Sch. admin.	.06**	.04*	-.05*		.03	.02	.04	-.04	-.03	-.01	.17**	.05*	-.04	.08**	-.02	.04*	.04	-.02	.08**	.02
5. Tch. gender	-.09**	.07**	.11**	.10**		-.15**	.13**	.04	.06**	.05*	-.02	.17**	-.12**	.05*	.00	.00	-.04	.00	.02	-.02
6. Tch. experience	-.09**	-.07**	-.05*	.19**	.00		-.05*	.14**	.03	-.03	.19**	.08**	.00	.18**	.00	.05*	.05*	.05*	.06**	.07**
7. Tch. number	-.06**	-.03	-.10**	.02	.06*	.01		.30**	.21**	-.07**	-.02	.00	-.03	.05**	-.04*	.09**	.08**	-.01	.05*	.02
8. Tch. geometry	.08**	.04	-.03	.01	-.06**	-.11**	.39**		.47**	.06**	.02	.01	.13**	.02	.00	.04	.04	-.03	.02	.00
9. Tch. data	.08**	.03	-.15**	.02	.00	.10**	.50**	.50**		.02	.03	.00	.06**	.11**	-.01	.03	.03	-.01	-.03	.04*
10. Tch. hw.	-.06*	-.06*	-.01	-.06**	.01	.02	.05*	-.06**	.05*		.06**	.02	-.02	.00	.02	.04	.00	.00	.06**	.03
11. Tch. attitudes	.02	.05*	.13**	-.02	.02	.13**	-.02	-.03	.04	.19**		.32**	.22**	.22**	.00	.11**	.06**	.03	.07**	.04
12. Cla. homo.	-.23**	-.14**	.12**	.01	-.08**	.03	-.01	.02	-.01	.04	.23**		.22**	.28**	.01	.17**	.10**	.06**	.12**	.03
13. Cla. physi. res.	.19**	.05*	.02	.06**	.03	-.04	-.08**	-.04	.00	.06*	.26**	.11**		.34**	.00	.01	.02	.04	.01	-.01
14. Cla. Instru. res.	.12**	.11**	.03	-.08**	-.10**	-.18**	-.14**	.05*	.01	.11**	.28**	.26**	.35**		.00	.02	.03	.03	.00	.11**
15 Stu. gender.	.01	.01	-.03	.03	.00	.06**	.00	-.06*	-.03	-.03	-.04	-.03	-.04	-.05*		-.01	.05*	.00	.08**	.05*
16 Stu. mo. edu.	-.10**	.02	.07**	-.02	.01	.08**	.00	.00	-.04	.05*	.10**	.02	-.02	.03	-.05*		.30**	.10**	.26**	.04
17 Stu. books	-.02	.02	.04	-.03	-.08**	.06**	.05*	.05*	.08**	.01	.10**	.03	.00	.09**	.01	.29**		.13**	.20**	.06**
18 Stu. motiv.	-.03	.02	-.03	.02	.00	.05*	.02	.05	.05*	.06*	.00	.02	-.02	.05	-.06**	.10**	.09**		.28**	.05*
19 Stu. edu. exp.	-.05*	.07**	.00	-.03	-.01	.01	.01	.03	.00	.08**	.06**	.05	.00	.03	.09**	.30**	.17**	.22**		.06**
20 Stu. hw	-.04	-.03	.00	.01	-.04	-.03	.00	-.04	-.02	.06**	.01	.03	-.04	.07**	.09**	.00	.08**	.03	.04*	

The British Columbia correlation matrix is above the diagonal, and the Ontario matrix is below the diagonal. *$p \leq .05$, **$p \leq .01$.

All three methods were used to analyze each of the five plausible values in mathematics as the dependent variable for each province, British Columbia and Ontario. As stated in Chapter 3, plausible values were randomly drawn from the distribution of possible scores for the students assuming that they had answered all of the test items. All the plausible values shared the same mean but incorporated different error components. In the TIMSS data, because of the well designed test and large sample, the error component in each plausible value is actually small. Using an appropriate analysis method, the analysis of the five plausible values for students in British Columbia and Ontario should produce consistent results within each province because an appropriate analysis should be robust to relatively minor measurement errors. In other words, since the distributions across the plausible values are similar, the measurement errors of the independent variables should not affect the estimated parameters.

Another factor to consider when comparing the results of the five plausible values is the accuracy of the estimated standard errors, which should have confidence intervals that are narrow and similar. Finally, all results should be reasonably consistent with the theoretical literature, which is reviewed in Chapter 3. When interpreting the results, only regression coefficients statistically significant and above ±.05 are discussed. For these analyses, a rule of thumb is that coefficients between -.10 and +.10 are small effects; ±.10– ±.19 are moderate; coefficients ±.20 and above are large. For example, a one standard deviation in an independent variable must result in 10% of a standard deviation change (+.10 or -.10) in the dependent variable to be discussed as a small effect.

Table 43
Aggregate Analysis in British Columbia

	PVMATH1			PVMATH2			PVMATH3			PVMATH4			PVMATH5		
	Unstandardized			Unstandardized			Unstandardized			Unstandardized			Unstandardized		
	B	SE	Beta	B	SE	Beta	B	SE	Beta	B	SE	Beta	B	SE	Beta
Constant	510.02	2.86		510.57	2.90		510.72	3.01		510.65	2.89		511.79	2.80	
School															
Physical resources	-1.10	1.70	-.05	-1.33	1.73	-.06	-1.81	1.79	-.08	-1.82	1.72	-.09	-1.72	1.66	-.08
Instruc. resources	-.73	.81	-.07	-.57	.83	-.06	-.62	.85	-.06	-.30	.82	-.03	-.55	.80	-.06
Average SES	3.77	4.25	.08	1.98	4.32	.04	1.76	4.47	.04	1.35	4.29	.03	2.35	4.16	.05
Administration	.12	.18	.05	.08	.18	.04	.05	.19	.02	.08	.18	.03	.06	.17	.03
Teacher/classroom															
Gender	3.42	7.15	.04	4.44	7.26	.05	5.12	7.51	.05	6.13	7.22	.06	5.06	7.00	.06
Experience	1.56	1.82	.06	1.70	1.85	.07	2.19	1.91	.09	1.60	1.84	.06	1.47	1.78	.06
Number and algebra	1.39	.71	.16	1.31	.72	.15	1.16	.74	.13	1.25	.72	.14	1.22	.69	.14
Geometry	.76	.64	.10	.67	.65	.08	.76	.68	.10	.63	.65	.08	.64	.63	.08
Data and chance	.25	1.01	.02	.87	1.03	.07	.35	1.06	.03	.46	1.02	.04	.51	.99	.04
Assign. of homework	1.51	3.46	.03	1.80	3.51	.04	1.95	3.64	.04	1.95	3.50	.04	.79	3.39	.02
Attitudes	-1.95	1.93	-.08	-2.02	1.96	-.08	-1.99	2.03	-.08	-2.01	1.95	-.08	-2.10	1.89	-.09
Homo. of students	4.14***	1.10	.31	3.78***	1.11	.28	3.42**	1.15	.25	3.70***	1.11	.28	3.91***	1.07	.30
Physical resources	-1.59	2.49	-.05	-.07	2.53	.00	-.73	2.62	-.02	-.61	2.52	-.02	-.41	2.44	-.01
Instruc. resources	.06	1.46	.00	-.16	1.48	-.01	.14	1.54	.01	-.22	1.48	-.01	-.69	1.43	-.04
Student															
Gender	-28.76	23.07	-.09	-26.53	23.41	-.08	-37.95	24.22	-.12	-37.05	23.28	-.12	-38.16	22.56	-.12
Mother education	3.80	5.78	.07	4.34	5.87	.08	6.12	6.07	.11	5.85	5.84	.10	3.43	5.66	.06
Number of books	13.70	7.52	.16	15.57*	7.63	.18	13.77	7.89	.16	17.62*	7.59	.20	13.89	7.35	.17
Instru. motivation	2.42	4.36	.05	2.06	4.43	.04	1.14	4.58	.02	1.22	4.40	.02	1.02	4.27	.02
Edu. expectations	23.71***	6.60	.36	24.10***	6.70	.37	25.48***	6.93	.38	22.53***	6.66	.34	24.99***	6.46	.39
Effort in homework	-2.12	9.87	-.02	-.83	10.02	-.01	-2.56	10.37	-.02	3.04	9.97	.02	5.31	9.66	.04
R^2		.57			.56			.54			.56			.56	

*$p \leq .05$, **$p \leq .01$, ***$p \leq .001$.

Table 44

Aggregate Analysis in Ontario

	PVMATH1			PVMATH2			PVMATH3			PVMATH4			PVMATH5		
	Unstandardized			Unstandardized			Unstandardized			Unstandardized			Unstandardized		
	B	SE	Beta	B	SE	Beta	B	SE	Beta	B	SE	Beta	B	SE	Beta
Constant	519.19	2.41		519.77	2.19		519.34	2.36		519.11	2.14		519.10	2.34	
School															
Physical resources	-1.77	1.30	-.11	-1.69	1.18	-.11	-1.40	1.28	-.09	-2.44*	1.16	-.16	-1.37	1.26	-.09
Instruc. resources	.99	.70	.11	1.03	.64	.13	.48	.69	.06	.86	.63	.10	.55	.68	.06
Average SES	4.82	3.03	.12	5.00	2.75	.13	6.18*	2.98	.16	6.07*	2.70	.16	5.84*	2.95	.15
Administration	-.04	.15	-.02	-.07	.14	-.04	-.06	.15	-.03	-.08	.13	-.04	-.01	.15	-.01
Teacher/classroom															
Gender	2.08	5.39	.03	1.86	4.90	.03	-.71	5.30	-.01	1.94	4.80	.03	.83	5.24	.01
Experience	.64	1.62	.03	2.04	1.47	.10	1.00	1.59	.05	.86	1.44	.04	1.05	1.58	.05
Number and algebra	.91	.49	.15	.94*	.45	.17	1.04*	.49	.18	.72	.44	.13	.89	.48	.15
Geometry	-.10	.62	-.01	.22	.56	.03	-.13	.61	-.02	.10	.55	.01	-.08	.60	-.01
Data and chance	.15	1.07	.01	-.77	.98	-.07	-.78	1.06	-.07	.12	.96	.01	-.33	1.04	-.03
Assign. of homework	4.20	2.96	.11	3.40	2.69	.09	2.54	2.91	.07	4.15	2.63	.11	3.56	2.88	.09
Attitudes	.36	1.40	.02	.47	1.27	.03	.48	1.37	.03	.28	1.24	.02	.62	1.36	.04
Homo. of students	.40	.81	.04	-.21	.74	-.02	-.17	.80	-.02	-.07	.72	-.01	.25	.79	.02
Physical resources	-2.90	1.77	-.13	-2.13	1.61	-.10	-2.33	1.74	-.11	-2.37	1.57	-.11	-2.98	1.72	-.13
Instruc. resources	1.54	.93	.13	1.68*	.85	.16	1.57	.92	.14	1.24	.83	.11	1.41	.91	.12
Student															
Gender	-9.07	13.33	-.05	-8.03	12.11	-.05	-6.95	13.09	-.04	.64	11.86	.00	-4.77	12.96	-.03
Mother education	7.25	3.78	.17	7.97*	3.43	.20	8.38*	3.71	.20	9.64**	3.36	.24	12.08**	3.67	.28
Number of books	15.33***	4.70	.28	13.12**	4.27	.26	14.55**	4.62	.27	12.23**	4.18	.24	11.49*	4.57	.21
Instru. motivation	1.59	3.11	.04	2.93	2.83	.08	2.00	3.06	.05	3.31	2.77	.09	-.57	3.02	-.02
Edu. expectations	3.95	4.34	.08	3.94	3.94	.08	5.89	4.27	.12	6.19	3.86	.13	6.19	4.22	.12
Effort in homework	-15.52**	6.06	-.20	-8.35	5.50	-.11	-12.88*	5.95	-.16	-12.03*	5.40	-.16	-15.33**	5.89	-.19
R^2		.37			.41			.38			.45			.40	

*$p \leq .05$, **$p \leq .01$, ***$p \leq .001$.

Aggregate Analysis

In British Columbia, the 2,392 students remaining after the preliminary analyses have been aggregated into 134 schools, and in Ontario, the 2,080 students have been aggregated into 169 schools. As such, the sample size for the analysis was 134 schools in British Columbia and 169 schools in Ontario. Table 43 reports the constants, the unstandardized regression coefficients and their standard errors, the standardized regression coefficients, and the R^2s of the aggregate analysis for British Columbia. Table 44 reports the results of the same analysis for Ontario. In these tables, the student-level variables, which are characteristics that students bring to schools from their homes, are listed at the bottom. The school and classroom/teacher level variables, the characteristics of the schools, are listed above the student variables.

In the five sets of analysis in British Columbia for each of the plausible values of the mathematics achievement, the constant was relatively stable (510.02–511.79) and so were the standard errors (2.80–3.01). Consequently, the 95% confidence interval for the students' average scores in mathematics ranged from 504.41 to 517.28. Additionally, the percentages of variance explained for the plausible values were relatively stable (R^2=.54–.57), but these coefficients were much higher than those usually found in the literature (Reynolds, et. al., 2000; Townsend, 2007). They are likely to be inflated from aggregating the data (Robinson, 1950).

Among all the independent variables, the largest and a strong effect was from the students' educational expectation (B=.34–.39, p<.001), which means that one standard deviation increase in the students' average educational expectations is associated with a 34 to 39 percent standard deviation increase in the students' aggregated scores in mathematics. This result is supported by the literature reviewed in Chapter 3 that the students' educational expectations has a strong impact on their achievement (Dweck & Sorich, 1999; Kaplan & Maehr, 1999; Marjoribanks, 2005; Roeser, Midgely, & Urdan 1996). The next largest effect was from the homogeneity of students in classroom ($B=.25–.31$, $p<.001$), which means that the more homogeneous the students are in a classroom, the higher their average achievement in mathematics. This is an effect that is also supported by the literature (Kulik, 1992; Kulik & Kulik, 1982). Finally, the number of books in home had a significant and moderately strong effect ($B=.18$, $.20$, $p<.05$) for two plausible values, PVMATH2 and PVMATH4. The fact that this variable did not have significant effects for the other plausible values is surprising ($B=.16$, $.17$, $p>.05$). In these analyses the mother's education did not have a significant effect on the students' achievements, which should not be surprising because students with a large amount of missing values on their mothers' education, whose mothers probably have received less education than the remaining students' mothers, were excluded from this study. Consequently, the explanatory power of this variable has been reduced due to its restricted variance. However, these results probably indicate that the number of books in home is more important than the mother's education because this variable reflects how much the family values reading, specifically, and education, generally.

Similarly, in Ontario, the constants were relatively stable and higher than those in British Columbia (519.10–519.77) and their standard errors were also stable and slightly lower (2.14–2.41) probably because the sample size was larger in Ontario. Consequently, the 95% confidence interval for the students' average scores in mathematics ranged from 514.47 to 524.06, higher than the confidence interval in British Columbia. However, the

R^2 was lower and varied more (R^2=.37–.45) than those in British Columbia. Among all the independent variables, there were more significant effects in Ontario, but there were fewer consistencies across the five plausible values. Similar to British Columbia, the number of books in home had the largest effect and it was consistently significant, but the coefficients varied considerably (B=.21–.28, p<.001). The effects of the other variables were even less consistent, and for some plausible values they were statistically significant but not significant for other plausible values. Mother's education had statistically significant effects for four plausible values (B=.20 –.28, p<.05 or p<.01) but not for the other value (B=.17), which was different from British Columbia where mother's education did not affect the students' achievement. Effort in completing homework was also significant for four plausible values, but the effects were negative (B= -.16 – -.20, p<.05 and p<.01). The fact that the effort in completing homework was negatively related to the students' achievement in mathematics probably indicates that the less able students need more time to complete their homework. The average SES of students had statistically significant effects for three plausible values (B=.15 –.16, p<.05), but not for the other two values (B=.12, .13, p>.05). According to the literature, the SES of schools often has a strong effect on the students' aggregated achievement. Teaching numbers and algebra had significant effects for two plausible values (B=.17, .18, p<.05) but not for the other three plausible values (B=.13–.15). And, the physical resources in school had a significant negative effect for one plausible value (B= -.16, p<.05) but not for the other four values (B= -.09– -.11, p>.05). This negative effect suggests that students do less well in schools with an abundance of physical resources than in schools with fewer resources, which is not consistent with many other studies (Cash, 1993; Earthman, 2004; Earthman

& Lemasters, 1996; Heschong Mahone Group, 1999; McGuffey & Brown, 1978; Tanner, 2000; Tanner & Lackney, 2006).

Overall, the results reported for British Columbia in Table 43 and for Ontario in Table 44 suggest that the two provinces had considerable differences. Students in Ontario had higher average scores than students in British Columbia, but the analysis in Ontario explained considerably less variance. There were also inconsistent estimates for the independent variables, which probably did not reflect realistic parameters, especially in Ontario. It is possible that there was considerably more variation in the students within schools in Ontario than in British Columbia, but it was difficult to test this assumption with an aggregate analysis of these data. Another factor that may cause less consistency in Ontario is that the school sizes were more balanced in British Columbia than in Ontario. In the next section, the results of the disaggregated analysis are reported.

Disaggregate Analysis

In the disaggregate analysis, the school level variables have been distributed to the student level and the student was the unit of analysis. Thus, the sample size was 2,392 grade eight students in British Columbia and 2,080 grade eight students in Ontario. Table 45 reports the constants, the unstandardized coefficients and their standard errors, the standardized coefficients, and the R^2s for the five plausible values in disaggregate analysis for British Columbia. Table 46 reports the results of the same analyses for Ontario.

Table 45

Disaggregate Analysis in British Columbia

	PVMATH1			PVMATH2			PVMATH3			PVMATH4			PVMATH5		
	Unstandardized			Unstandardized			Unstandardized			Unstandardized			Unstandardized		
	B	SE	Beta	B	SE	Beta	B	SE	Beta	B	SE	Beta	B	SE	Beta
Constant	517.05	1.97		519.02	2.02		519.38	2.04		517.99	1.99		519.28	1.99	
School															
Physical resources	-.33	.81	-.01	-.68	.83	-.02	-1.03	.84	-.03	-1.39	.82	-.04	-1.01	.81	-.03
Instruc. resources	-.08	.40	-.01	.08	.41	.00	.15	.41	.01	.41	.40	.02	.06	.40	.00
Average SES	2.54	1.75	.03	1.09	1.79	.01	1.83	1.81	.02	2.05	1.77	.03	1.62	1.76	.02
Administration	.24**	.08	.06	.22**	.08	.06	.24**	.08	.06	.23**	.08	.06	.22**	.08	.06
Teacher/classroom															
Gender	-1.61	2.98	-.01	-1.61	3.05	-.01	-.41	3.08	.00	.71	3.01	.01	-.54	3.01	.00
Experience	1.55*	.80	.04	1.72*	.82	.04	2.24**	.83	.06	1.46	.81	.04	1.53	.81	.04
Number and algebra	2.00***	.27	.16	2.01***	.28	.16	1.97***	.28	.15	1.99***	.28	.16	1.91***	.28	.15
Geometry	.64*	.29	.05	.62*	.29	.05	.71*	.30	.06	.56*	.29	.05	.51	.29	.04
Data and chance	-.61	.42	-.03	-.15	.43	-.01	-.56	.43	-.03	-.28	.42	-.02	-.47	.42	-.03
Assign. of homework	2.89*	1.50	.04	3.67*	1.54	.05	3.34*	1.55	.04	3.51*	1.51	.05	2.62	1.51	.04
Attitudes	-1.43	.84	-.04	-1.68*	.86	-.04	-1.63	.87	-.04	-1.54	.85	-.04	-1.73*	.84	-.05
Homo. of students	4.97***	.47	.24	4.93***	.48	.23	4.57***	.48	.21	4.65***	.47	.22	4.90***	.47	.23
Physical resources	-1.07	1.04	-.02	.12	1.06	.00	-.20	1.07	.00	-.20	1.05	.00	-.43	1.05	-.01
Instruc. resources	-.56	.61	-.02	-.83	.63	-.03	-.73	.63	-.03	-.35	.62	-.01	-.84	.62	-.03
Student															
Gender	-3.03	2.73	-.02	-7.07*	2.80	-.05	-7.24**	2.82	-.05	-3.99	2.76	-.03	-5.93*	2.75	-.04
Mother education	1.19	.84	.03	1.46	.86	.04	1.18	.87	.03	1.72*	.85	.04	1.40	.85	.04
Number of books	9.60***	1.18	.17	10.54***	1.21	.18	10.04***	1.22	.17	10.25***	1.19	.18	10.56***	1.19	.19
Instru. motivation	3.76***	.62	.12	3.66***	.64	.12	3.33***	.64	.11	3.23***	.63	.11	3.40***	.63	.11
Edu. expectations	7.78***	.89	.19	7.54***	.91	.18	8.58***	.92	.20	8.05***	.90	.19	7.84***	.90	.19
Effort in homework	-10.05***	1.56	-.13	-10.80***	1.59	-.13	-10.57***	1.61	-.13	-10.31***	1.57	-.13	-9.96***	1.57	-.13
R^2		.25			.25			.24			.25			.25	

*$p \leq .05$, **$p \leq .01$, ***$p \leq .001$.

Table 46

Disaggregate Analysis in Ontario

	PVMATH1			PVMATH2			PVMATH3			PVMATH4			PVMATH5		
	Unstandardized			Unstandardized			Unstandardized			Unstandardized			Unstandardized		
	B	SE	Beta	B	SE	Beta	B	SE	Beta	B	SE	Beta	B	SE	Beta
Constant	522.58	2.84		525.50	2.91		524.99	2.88		521.33	2.90		524.53	2.88	
School															
Physical resources	-2.98***	.73	-.10	-3.15***	.75	-.11	-2.73***	.74	-.10	-3.62***	.75	-.13	-2.82***	.74	-.10
Instruc. resources	1.03*	.41	.06	1.15**	.42	.07	.74	.41	.04	1.09**	.42	.06	.77	.41	.05
Average SES	6.42***	1.80	.08	6.94***	1.84	.09	8.40***	1.83	.11	7.08***	1.84	.09	7.17***	1.83	.09
Administration	.05	.09	.01	-.01	.09	.00	.01	.09	.00	-.01	.09	.00	.04	.09	.01
Teacher/classroom															
Gender	2.98	3.03	.02	.61	3.09	.00	.18	3.07	.00	2.56	3.09	.02	1.49	3.06	.01
Experience	.49	.88	.01	1.56	.90	.04	1.00	.89	.03	1.05	.90	.03	.87	.89	.02
Number and algebra	.38	.31	.03	.55	.31	.05	.58	.31	.05	.36	.31	.03	.56	.31	.05
Geometry	1.00**	.36	.07	.91*	.37	.07	.84*	.36	.06	.96**	.37	.07	1.03**	.36	.08
Data and chance	-.11	.65	-.01	-.62	.66	-.03	-.86	.65	-.04	-.09	.66	.00	-.62	.65	-.03
Assign. of homework	4.24**	1.57	.06	3.67*	1.60	.05	2.91	1.59	.04	4.63**	1.60	.07	2.86	1.59	.04
Attitudes	.20	.76	.01	.26	.77	.01	.40	.77	.01	.04	.77	.00	.40	.76	.01
Homo. of students	-.24	.48	-.01	-.63	.49	-.03	-.58	.49	-.03	-.39	.49	-.02	-.34	.49	-.02
Physical resources	-.91	.95	-.02	-1.05	.97	-.03	-.95	.96	-.02	-.85	.97	-.02	-1.26	.96	-.03
Instruc. resources	1.23*	.55	.06	1.43**	.56	.07	1.49**	.55	.07	1.17*	.56	.05	1.27*	.55	.06
Student															
Gender	-9.09**	2.92	-.07	-10.37***	2.98	-.08	-8.33**	2.96	-.06	-7.11*	2.98	-.05	-9.78***	2.95	-.07
Mother education	2.92**	.93	.08	2.95**	.95	.08	3.21***	.94	.08	2.41**	.95	.06	3.74***	.94	.10
Number of books	13.49***	1.25	.25	12.39***	1.28	.23	12.67***	1.27	.23	12.98***	1.28	.24	12.13***	1.27	.22
Instru. motivation	3.40***	.71	.11	3.07***	.72	.10	3.34***	.72	.11	3.23***	.72	.10	3.24***	.72	.10
Edu. expectations	7.33***	1.00	.17	7.06***	1.03	.17	7.26***	1.02	.17	7.70***	1.03	.18	7.39***	1.02	.17
Effort in homework	-12.07***	1.69	-.16	-10.93***	1.73	-.14	-12.37***	1.71	-.16	-11.90***	1.72	-.15	-12.74***	1.71	-.17
R^2		.23			.21			.22			.22			.22	

**p* ≤ .05, ***p* ≤ .01, ****p* ≤ .001.

Overall, the constant for each plausible value for the students' mathematics achievement in British Columbia, was stable across the five analyses (517.05–519.38), which was higher than in the aggregate analysis (510.02–511.79), and, as expected, the standard errors (1.97–2.04) were smaller than in the aggregate analysis (2.80–3.01). Additionally, the percentages of variance explained for each of the plausible values were relatively stable and reasonable (R^2=.24–.25) according to the explained variance generally found in the research literature (Reynolds, et. al., 2000; Townsend, 2007).

Obviously, there were more significant effects among the independent variables in the disaggregate analysis than in the aggregate analysis simply because the number of observations was much larger. In British Colombia, almost all the student variables had statistically significant effects except for the mother's education variable. More specifically, the students' educational expectations had the largest effect and was moderate (B=.18–.20, p<.001), which means that a one standard deviation increase in the students' educational expectation is associated with between 18% and 20% of a standard deviation increase in their mathematics score. In the aggregate analysis for British Columbia, the students' educational expectations variable had an effect that was almost twice the size of the coefficients in the disaggregate analysis. The next largest effect was the number of books in home (B=.17–.19, p<.001), which was statistically significant for all the plausible values. Surprisingly, these coefficients were similar to the coefficients in the aggregate analysis (B=.16–.20, p<.05 or p>.05). The effort in completing homework had a statistically significant and modest effect (B= -.13, p<.001) for all plausible values. Finally, the students' instructional motivation also had a statistically significant and modest effect (B=.11–.12, p<.001) on their achievement, which only had a small and non-significant effect in the aggregate analysis. All these effects are supported by the research literature reviewed in Chapter 3. Consistent with the literature, the students' gender had a small and negative effect favoring the males ($B = -.02 – -.05$, $p<.05$ or $p<.01$), which was statistically significant for three of the five plausible values.

Among the school and teacher/classroom variables, the largest effect was from the homogeneity of students in classroom ($B=.21–.24$, $p<.001$), which was similar to, but smaller than, the result of this variable in the aggregate analysis. This actually shows, at the student level, that the more similar the students' classmates are, the higher their scores in mathematics. The teaching number and algebra variable had the next largest effect, but it was moderate ($B=.15–.16$, $p<.001$), which indicates that a one standard deviation increase in the instructional time teachers spend on teaching number and algebra increases their students' achievement in mathematics by between 15 and 16 percent of a standard deviation. Simply put, this effect indicates that teachers who spend more time teaching a subject have students that understand the subject better, which is not surprising and is supported by the research literature (Borg, 1979; Brophy & Evertson, 1976; Cooley & Leinhardt, 1980; Dunkin & Doenau, 1980; Good & Becherman,1978; Good, Grouws, & Beckerman, 1978; Wyne & Stuck, 1982). This effect has also been found in the aggregate analysis, but none of the five coefficients were statistically significant. It is interesting that the instructional time teachers spent on teaching geometry only had a small effect and it was statistically significant for only four plausible values ($B=.05–.06$, $p<.05$). The other standardized coefficients were well below $\pm.05$, thus they are not discussed.

Likewise, in Ontario the constants were relative stable (521.33– 525.50), higher than those in the aggregate analysis, and their standard errors were also stable (2.84–2.91) and slightly larger than in the aggregate analysis. The R^2 was relatively stable and reasonable (R^2=.21–.23), but slightly smaller than in British Columbia (R^2=.24–.25), which means that the model explains less variance in Ontario.

As well, there were more significant effects among the independent variables in the disaggregate analysis than in the aggregate analysis, and there was more consistency in the coefficients across the five plausible values. Similar to British Columbia, all the student variables had statistically significant effects and they appeared to be more consistent than those in British Columbia. The number of books in home had the largest effect (B=.22–.25, p<.001), which was slightly smaller than the one in the aggregate analysis. The students' educational expectations variable had the second largest but moderate effect (B=.17–.18, p<.001), which was very similar to the coefficients in British Columbia. Effort in completing homework also had a moderate but negative effect (B= -.14– -.17, p<.001), also similar to the coefficients in British Columbia. The instructional motivation of the students had a statistically significant but moderate effect (B=.10–.11, p<.001), and the mother's educational level had a statistically significant and small to moderate effect (B=.06–.10, p<.001 or p<.01) meaning that the more years of education that mothers received, the higher the scores their children achieve in mathematics. Similar to the coefficients in the disaggregate analysis for British Columbia, the students' gender had a small effect favouring male students, but this effect was stronger and more consistent in the disaggregate analysis.

Among the school variables, the physical resources in school had a moderate but negative effect (B= -.10– -.13, p<.001), which was similar to the result in the aggregate analysis. This may look surprising because it is not consistent with many findings in many previous studies, which suggested that students did better in schools with more resources and not in schools with fewer resources (Berner, 1993; Buckley, Schneider, & Shang, 2004; Cash, 1993; Earthaman, 2004; Lewis, 2000; Roberts, 2009; Tanner, 2000; Tanner & Lackney, 2006; Uline & Tschannen-Moran, 2008; Uline, Tschannen-Moran, & Wolsey, 2009). Next, the average SES of the schools had a statistically significant but only a small to moderate effect (B=.08–.11, p<.001), similar to the aggregate analysis. This effect is consistent with the research literature (Alexander & Eckland, 1975; Alwin & Otto, 1977; Bain & Anderson, 1974; Bidwell, 1972; Haller & Woelfel, 1972; Hauser, Sewell, & Alwin, 1976; Meyer, 1970; Nelson, 1972; Willms & Smith, 2004; Woelfel & Haller, 1971). Unlike in the aggregate analysis, the instructional time on teaching number and algebra did not have a significant effect. Instead, the instructional time on teaching geometry had a small and statistically significant effect (B=.06–.08, p<.05 or p<.01). Finally, the other coefficients were small and are not discussed.

Overall, the results reported for British Columbia in Table 45 and Ontario in Table 46 confirmed that the two provinces had considerable differences in achievement scores of mathematics. More specifically, students in Ontario had higher average scores than students in British Columbia, but the confidence intervals overlapped considerably. In total, the analyses in Ontario explained slightly less variance than the analyses in British Columbia. Finally, the disaggregate analyses presented more consistent results than the aggregate analyses. In essence, the effects of schools, teachers/classrooms, and students themselves on their achievement in mathematics were different in the two provinces. Compared to the aggregate analyses, the regression coefficients in disaggregate analysis were more stable and more consistent with the research literature. Nevertheless, the results may be biased due to the violation of assumptions disaggregate analysis made about the data at the student level. Consequently, there was greater consistency in the results for Ontario than for British Columbia probably because the clustering effect was smaller in Ontario due to a larger school sample and a smaller student sample in each school. Of course, it was difficult to test this assumption in a disaggregate analysis. In the next section, the results of the HLM analysis are reported; the assumption is that HLM improves on the parameter estimates explaining achievement scores of mathematics in both British Columbia and Ontario.

HLM Analysis

In the HLM analysis, the school and classroom/teacher variables and the student variables were analyzed at two levels. Thus, the samples were 2,392 grade eight students at a lower level and 134 of schools at a higher level in British Columbia, and 2,080 grade eight students at a lower level and 169 of schools at a higher level in Ontario. Table 47 reports the constants, the unstandardized coefficients and their standard errors, the standardized coefficients, and the percentages of variance explained for the five plausible values for British Columbia and Table 48 reports the same results for Ontario.

Table 47

HLM Analysis in British Columbia

	PVMATH1			PVMATH2			PVMATH3			PVMATH4			PVMATH5		
	Unstandardized			Unstandardized			Unstandardized			Unstandardized			Unstandardized		
	B	SE	Beta	B	SE	Beta	B	SE	Beta	B	SE	Beta	B	SE	Beta
Constant	512.29	3.17		515.87	3.34		515.68	3.34		513.34	3.18		515.92	3.25	
School															
Physical resources	-.88	1.35	-.02	-1.27	1.49	-.03	-1.63	1.42	-.01	-1.72	1.43	-.05	-1.44	1.43	-.04
Instruc. resources	-.06	.90	-.01	.14	.92	.00	-.19	.93	-.04	.39	.90	.01	.12	.94	.00
Average SES	6.19	4.47	.07	4.97	4.78	.05	4.70	5.02	.05	3.94	4.49	.06	4.20	4.77	.05
Administration	.34*	.15	.08	.33*	.15	.07	.28	.16	.07	.31*	.15	.07	.27	.15	.07
Teacher/classroom															
Gender	2.99	7.47	.01	4.09	7.72	.02	5.12	7.65	.02	4.21	7.43	.02	4.09	7.68	.02
Experience	2.92	1.65	.07	2.91	1.72	.07	3.57	1.80	.09	2.94	1.63	.07	2.86	1.64	.07
Number and algebra	2.39***	.54	.16	2.40***	.56	.15	2.33***	.55	.15	2.46***	.56	.16	2.35***	.56	.15
Geometry	.37	.64	.05	.23	.65	.04	.27	.64	.05	.13	.64	.04	.19	.66	.03
Data and chance	-.13	.92	-.01	.40	.93	.02	-.06	.95	.00	.11	.92	.01	.12	.93	.01
Assign. of homework	5.40	3.47	.05	6.03	3.70	.05	6.38	3.78	.05	6.01	3.56	.06	5.29	3.66	.04
Attitudes	-2.82	1.68	-.05	-3.13	1.74	-.05	-2.90	1.73	-.05	-3.10	1.66	-.05	-2.98	1.64	-.05
Homo. of students	5.51***	1.21	.25	5.31***	1.24	.23	5.12***	1.33	.22	5.38***	1.20	.24	5.28***	1.22	.24
Physical resources	-1.19	2.66	-.02	.43	2.79	.01	-.39	2.84	.00	-.20	2.69	.00	-.29	2.85	.00
Instruc. resources	-.25	1.38	-.03	-.40	1.47	-.04	-.31	1.41	-.03	-.33	1.38	-.03	-.87	1.52	-.05
Student															
Gender	-2.67	2.34	-.02	-7.38**	2.44	-.05	-7.00**	2.47	-.04	-3.42	2.43	-.02	-5.62*	2.55	-.04
Mother education	.60	.84	.01	.58	.81	.02	.40	.87	.01	.49	.84	.02	.42	.91	.02
Number of books	8.86***	.94	.16	10.06***	1.00	.17	9.64***	1.05	.16	9.61***	1.00	.16	10.06***	1.07	.17
Instru. motivation	3.62***	.57	.13	3.44***	.62	.12	3.49***	.60	.11	3.15***	.61	.11	3.42***	.60	.12
Edu. expectations	6.26***	.89	.16	6.14***	.89	.15	7.17***	.86	.17	6.64***	.85	.16	6.41***	.88	.16
Effort in homework	-11.17***	1.57	-.13	-11.94***	1.66	-.14	-11.31***	1.67	-.13	-11.38***	1.70	-.13	-11.36***	1.66	-.13
% of variance exp.	.26			.24			.24			.25			.23		

$^{*}p \leq .05$, $^{**}p \leq .01$, $^{***}p \leq .001$.

Table 48

HLM Analysis in Ontario

	PVMATH1			PVMATH2			PVMATH3			PVMATH4			PVMATH5		
	Unstandardized			Unstandardized			Unstandardized			Unstandardized			Unstandardized		
	B	SE	Beta	B	SE	Beta	B	SE	Beta	B	SE	Beta	B	SE	Beta
Constant	529.88	2.58		532.56	2.69		531.60	2.83		530.38	2.68		530.36	2.70	
School															
Physical resources	-3.79***	1.14	-.12	-3.75***	1.25	-.12	-3.33**	1.22	-.11	-4.55***	1.20	-.15	-3.66**	1.15	-.12
Instruc. resources	1.42*	.65	.09	1.28	.62	.08	1.04	.66	.06	1.32	.61	.08	1.04	.65	.06
Average SES	9.78***	2.88	.13	9.27**	2.72	.12	11.82***	2.83	.15	10.87***	2.68	.14	10.51***	2.98	.13
Administration	-.02*	.16	.00	-.09*	.16	-.02	-.04	.16	-.01	-.08*	.15	-.02	.02	.16	.01
Teacher/classroom															
Gender	9.73*	4.34	.07	6.66	4.72	.05	7.39	4.79	.05	8.50*	4.32	.06	6.06	4.59	.04
Experience	.48	1.31	.01	2.60*	1.24	.07	1.36	1.29	.03	1.23	1.31	.03	.10	1.35	.00
Number and algebra	.06	.44	.01	.16	.44	.01	-.08	.45	-.01	-.37	.41	-.03	-.07	.44	-.01
Geometry	.70	.49	.05	.80	.49	.06	.58	.54	.04	.82	.46	.06	.74	.53	.05
Data and chance	.40	1.18	.02	-.82	1.04	-.04	.00	1.16	.00	.29	1.08	.01	-.29	1.25	-.01
Assign. of homework	6.03*	2.35	.08	5.34*	2.24	.07	4.83*	2.39	.07	6.44**	2.34	.09	5.02*	2.38	.07
Attitudes	1.48	1.05	.05	1.88	1.11	.06	1.93	1.22	.06	1.80	1.11	.06	2.45*	1.22	.08
Homo. of students	.15	.71	.01	-.12	.65	-.01	-.11	.64	-.01	.06	.68	.00	-.04	.66	.00
Physical resources	-3.03*	1.38	-.07	-3.19*	1.37	-.08	-2.28	1.46	-.05	-3.58*	1.46	-.08	-3.37*	1.53	-.08
Instruc. resources	.18	1.01	.01	-.24	1.03	-.01	-.14	1.06	-.01	-.29	1.00	-.01	-.13	1.04	-.01
Student															
Gender	-8.69**	3.31	-.07	-11.74***	3.10	-.09	-9.43**	3.13	-.07	-7.98*	3.20	-.06	-9.45**	3.35	-.07
Mother education	1.63	1.11	.04	2.36*	1.03	.06	2.14*	1.07	.05	1.25	1.08	.03	2.64*	1.11	.07
Number of books	11.33***	1.44	.21	9.30***	1.45	.17	9.72***	1.40	.18	10.62***	1.45	.19	9.85***	1.40	.18
Instru. motivation	3.33***	.82	.11	2.91***	.78	.09	3.08***	.78	.10	2.96***	.84	.09	3.07***	.92	.10
Edu. expectations	7.70***	1.23	.18	7.68***	1.28	.18	7.48***	1.23	.17	7.84***	1.21	.18	7.44***	1.17	.17
Effort in homework	-15.82***	1.84	-.20	-14.63***	1.75	-.19	-16.19***	1.99	-.21	-15.34***	1.91	-.19	-16.64***	2.08	-.21
% of variance exp.	.21			.20			.21			.21			.20		

*$p \le .05$, **$p \le .01$, ***$p \le .001$.

In the analysis of British Columbia for each of the five plausible values in mathematics, the constant was stable across them (512.29–515.92), which was only slightly lower than in the disaggregate analysis. The constants' standard errors were also stable (3.17–3.34) and slightly higher than in the aggregate analysis. Additionally, the percentage of variance explained for each of the five plausible values was also stable and reasonable (.23–.26), similar to the R^2 in the disaggregate analysis.

Overall, in British Columbia, there were more significant effects among the independent variables in the HLM analysis than in the aggregate analysis. There were fewer significant effects in the HLM analysis than in the disaggregate analysis because the HLM analysis adjusts for group differences, and therefore, the standard errors are larger than in the disaggregate analysis. Similar to the disaggregate analysis, almost all the student variables had statistically significant effects except for the mother's education. More specifically, the number of books in the home had the largest effect and it was moderate (B=.16, .17, p<.001). The next largest effect on the plausible values was the student's educational expectations (B=.15–.17, p<.001). Among the school and teacher/classroom variables, the largest effect was from the homogeneity of students in classroom (B=.22–.25, p<.001), which was slightly larger than in the disaggregate analysis while much smaller than in the aggregate analysis. The next largest effect was from the instructional time the teachers spent on teaching number and algebra (B=.15, .16, p<.001), which was generally the same size as in the disaggregate analysis.

Similarly, in Ontario, the constant was stable across the five analyses (529.88–532.56), higher than in the disaggregate analysis and even higher than in the aggregate analysis. Their standard error was also stable (2.58–2.83) and higher than in the aggregate analysis, but smaller than in the disaggregate analysis. Additionally, the percentage of variance explained for each of the plausible values was stable and consistent (.20–.21) and only slightly smaller than in the disaggregate analysis. Also, in Ontario, there were more significant effects among the independent variables in the HLM analysis than in the aggregate analysis, but fewer significant effects than in the disaggregate analysis, especially for the school and teacher/classroom variables. More specifically, all student variable had a statistically significant effect and the coefficients were similar to the results of the disaggregate analysis. Particularly, the effort in completing homework had the strongest effect, and it was statistically significant but negative ($B= -.19– -.21$, $p<.001$), which means that a one standard deviation increase in the time students spent on completing their homework is associated with as almost one fifth standard deviation decrease in their scores in mathematics. The number of books in the home had the next strongest effect, which was statistically significant ($B=.17–.21$, $p<.001$). The educational expectations of students had a slightly smaller moderate yet statistically significant effect on achievement ($B=.17$, $.18$, $p<.001$). Finally, the students' instructional motivation had a statistically significant and moderate effect on achievement ($B=.09–.11$, $p<.001$). Among the school and teacher/classroom variables, the average SES of students in a school had the strongest effect and it was moderate and statistically significant ($B=.12–.15$, $p<.001$), which means that a one standard deviation increase in the average SES of the students in a school increases the average scores of the students by 12 to 15 percent of a standard deviation. The next strongest effect was based on the physical resources of the schools; the effect was statistically significant and negative ($B= -.11– -.15$, $p<.001$). Surprisingly, the physical resources in the classrooms also had a negative effect and four of the five coefficients were statistically significant (B= -.05– -.08, $p<.05$).

Similar to the aggregate and the disaggregate analyses, the results of the HLM analysis confirmed that there were considerable differences between British Columbia and Ontario. Specifically, students in Ontario had higher average scores than students in British Columbia, and the difference was larger in the HLM analysis than in either the aggregate or the disaggregate analysis. Additionally, the schools, teachers/classrooms, and students themselves had different effects on their achievement in mathematics. In total, the HLM analysis in Ontario explained slightly less variance than in British Columbia. As explained in Chapter 2, in HLM analysis produces more statistically significant effects among the student variables than the aggregate analysis, but fewer statistically significant effects for these variables than the disaggregate analysis. The next section summarizes the comparison between the three analytical methods so that the advantages of the HLM analysis are more clearly understood.

Comparing the Three Methods

As reported previously, all three analysis methods have produced relatively consistent constants and standard errors. However, there are some small but important differences between the three methods. More specifically, in both British Columbia and Ontario, the disaggregate and the HLM analyses both have slightly higher constants than those in the aggregate analysis partially because the gender variable has not been centered when entered into the disaggregate and HLM analyses. The standard errors of the constants are similar in the three analyses. In the disaggregate analysis for British

Columbia, the standard error of the constant is smaller than those in the aggregate and the HLM analyses. This observation may be due to the violation of the homogeneity assumption and the tendency of the disaggregate analysis to underestimate standard errors. In Ontario, the standard error of the constant in the disaggregate analysis is closer to the one in the HLM analysis because there is less clustering of students. It is noted that the standard error in the aggregate analysis is smaller in Ontario than in British Columbia partially because the school sample in Ontario is more than 25% larger (134 in British Columbia versus 169 schools in Ontario).

In addition, there are significant differences between the aggregate, the disaggregate, and the HLM analyses. Specifically, the percentages of variance explained by the three analysis methods are different. Overall, aggregate analysis appears to explain much higher percentages of variance for both British Columbia and Ontario than either disaggregate or HLM analysis. In fact, the R^2 in the aggregate analysis is extremely large, over .50, a value typically seen only in aggregate analysis in the research literature—a consequence of the decreased variance in the aggregate variables. Normally, the smaller the group size, the larger the R^2 value. In other words, the R^2 in an aggregate analysis shows the degree to which the data have been aggregated rather than the true explained variance. The aggregate analysis explains more variance in British Columbia than in Ontario simply because the school is the unit of analysis and the school sample size is smaller in British Columbia. The disaggregate and HLM analyses explain a similar amount of variance in the dependent variable, around .20. This is different from the result reported by Bryk and Raudenbush (1992), likely because they included several aggregate variables in their analysis.

Another significant difference among the three analytical methods is that in both British Columbia and Ontario, there are fewer significant coefficients in the aggregate analysis than in either the disaggregate or the HLM analysis. The number of statistically significant coefficients among the student variables are similar in the disaggregate and the HLM analyses. However, there are fewer statistically significant coefficients among the school level variables in the HLM than in the disaggregate analysis.

It is also noticeable that the three analytical methods have produced different regression coefficients and standard errors. Particularly, the coefficients and standard errors tend to vary more across the five plausible values in the aggregate analysis than in either the disaggregate or the HLM analysis. At the same time, the latter two methods produce similar consistencies in their estimates. To compare the differences in the coefficients and the standard errors among the three methods, Table 49 presents the summary of the results of the aggregate, the disaggregate, and the HLM analyses for plausible value one in British Columbia and Table 50 presents the same summary in Ontario. The results for one plausible value are presented, but the same trend is evident for the other four plausible values thus are not reported. Generally, in both British Columbia and Ontario, the parameters estimated in the disaggregate analysis are similar to those estimated in the HLM analysis in their magnitude, direction, and statistical significance, while those estimated in the aggregate analysis differ substantially from either of these two methods. In fact, the coefficients in the aggregated analysis should be questioned based on the literature reviewed in Chapter 3.

Table 49

Comparisons of Aggregate, Disaggregate, and HLM Analysis in British Columbia for Plausible Value One

	Aggregate			Disaggregate			HLM		
	B	SE	Beta	B	SE	Beta	B	SE	Beta
School variables									
Physical resources	-1.10	1.70	-.05	-.33	.81	-.01	-.88	1.35	-.02
Instruc. resources	-.73	.81	-.07	-.08	.40	-.01	-.06	.90	-.01
Average SES	3.77	4.25	.08	2.54	1.75	.03	6.19	4.47	.07
Administration	.12	.18	.05	.24**	.08	.06	.34*	.15	.08
Teacher variables									
Gender	3.42	7.15	.04	-1.61	2.98	-.01	2.99	7.47	.01
Experience	1.56	1.82	.06	1.55*	.80	.04	2.92	1.65	.07
Number and algebra	1.39	.71	.16	2.00***	.27	.16	2.39***	.54	.16
Geometry	.76	.64	.10	.64*	.29	.05	.37	.64	.05
Data and chance	.25	1.01	.02	-.61	.42	-.03	-.13	.92	-.01
Assign. of homework	1.51	3.46	.03	2.89*	1.50	.04	5.40	3.47	.05
Attitudes	-1.95	1.93	-.08	-1.43	.84	-.04	-2.82	1.68	-.05
Homo. of students	4.14***	1.10	.31	4.97***	.47	.24	5.51***	1.21	.25
Physical resources	-1.59	2.49	-.05	-1.07	1.04	-.02	-1.19	2.66	-.02
Instruc. resources	.06	1.46	.00	-.56	.61	-.02	-.25	1.38	-.03
Student variables									
Gender	-28.76	23.07	-.09	-3.03	2.73	-.02	-2.67	2.34	-.02
Mother education	3.80	5.78	.07	1.19	.84	.03	.60	.84	.01
Number of books	13.70	7.52	.16	9.60***	1.18	.17	8.86***	.94	.16
Instru. motivation	2.42	4.36	.05	3.76***	.62	.12	3.62***	.57	.13
Edu. expectations	23.71***	6.60	.36	7.78***	.89	.19	6.26***	.89	.16
Effort in homework	-2.12	9.87	-.02	-10.05***	1.56	-.13	-11.17***	1.57	-.13
R^2/**% of variance exp.**		.57			.25			.26	

**p* ≤ .05, ***p* ≤ .01, ****p* ≤ .001.

Table 50

Comparisons of Aggregate, Disaggregate, and HLM Analysis in Ontario for Plausible Value One

	Aggregate			Disaggregate			HLM		
	B	SE	Beta	B	SE	Beta	B	SE	Beta
School variables									
Physical resources	-1.77	1.30	-.11	-2.98***	.73	-.10	-3.79***	1.14	-.12
Instruc. resources	.99	.70	.11	1.03*	.41	.06	1.42*	.65	.09
Average SES	4.82	3.03	.12	6.42***	1.80	.08	9.78***	2.88	.13
Administration	-.04	.15	-.02	.05	.09	.01	-.02*	.16	.00
Teacher variables									
Gender	2.08	5.39	.03	2.98	3.03	.02	9.73*	4.34	.07
Experience	.64	1.62	.03	.49	.88	.01	.48	1.31	.01
Number and algebra	.91	.49	.15	.38	.31	.03	.06	.44	.01
Geometry	-.10	.62	-.01	1.00**	.36	.07	.70	.49	.05
Data and chance	.15	1.07	.01	-.11	.65	-.01	.40	1.18	.02
Assign. of homework	4.20	2.96	.11	4.24**	1.57	.06	6.03*	2.35	.08
Attitudes	.36	1.40	.02	.20	.76	.01	1.48	1.05	.05
Homo. of students	.40	.81	.04	-.24	.48	-.01	.15	.71	.01
Physical resources	-2.90	1.77	-.13	-.91	.95	-.02	-3.03*	1.38	-.07
Instru. resources	1.54	.93	.13	1.23*	.55	.06	.18	1.01	.01
Student variables									
Gender	-9.07	13.33	-.05	-9.09**	2.92	-.07	-8.69**	3.31	-.07
Mother education	7.25	3.78	.17	2.92**	.93	.08	1.63	1.11	.04
Number of books	15.33***	4.70	.28	13.49***	1.25	.25	11.33***	1.44	.21
Instru. motivation	1.59	3.11	.04	3.40***	.71	.11	3.33***	.82	.11
Edu. expectations	3.95	4.34	.08	7.33***	1.00	.17	7.70***	1.23	.18
Effort in homework	-15.52**	6.06	-.20	-12.07***	1.69	-.16	-15.82***	1.84	-.20
R^2/Variance explained		.37			.23			.21	

*$p \le .05$, **$p \le .01$, ***$p \le .001$.

Additionally, in both British Columbia and Ontario, among the student-level variables, the coefficients in the disaggregate and HLM analyses are very similar; they are slightly smaller in British Columbia than in Ontario. However, among the school-level variables, the discrepancy between the disaggregate and the HLM analyses is much larger in Ontario than in British Columbia, which is probably due to the larger schools and more balance school sample in British Columbia than in Ontario. This actually demonstrates the advantage of HLM analysis over disaggregate analysis when dealing with unbalanced groups because HLM, as noted previously, is capable of adjusting unbalanced groups (schools in this study) by borrowing information from other groups (Bryk & Raudenbush, 1992; Hox, 2002). On the contrary, disaggregate analysis cannot adequately compensate for the lack of information in extremely small groups, typically groups with less than five members, because there is not enough variance.

The discrepancy in the standard errors of the regression coefficients among the three analyses is also important. Overall, the estimates of the standard errors are consistently overestimated for the student-level variables in the aggregate analysis, and they are consistently underestimated for the school-level variables in the disaggregate analysis. These results were expected due the respective estimating procedures explained in Chapter 2. In comparison, the HLM analysis produces more accurate standard errors for the coefficients at both the school-level and the student-level. More specifically, for the student-level variables, the standard errors in the aggregate analysis are around seven times larger than those in the disaggregate analysis for British Columbia and four times larger for Ontario, which suggests that the size of schools can strongly affect the results at the individual level in the aggregate analysis. The standard errors in the disaggregate and the HLM analyses are very similar for both British Columbia and Ontario, but they are smaller in British Columbia (around 10% less), due to the 15% fewer total number of students sampled in this province. For the school-level variables, the standard errors in the aggregate and the HLM analyses are very similar, which are about twice as large as those in the disaggregate analysis for British Columbia and about one and a half times larger for Ontario, probably because of the violation of the homogeneity of data assumption in the disaggregate analysis. Nevertheless, the fact that the standard errors of the coefficients for the school-level variables in the aggregate analysis are more valid than in the disaggregate analysis is not meaningful because the coefficients themselves are probably not valid as based on the literature reviewed in Chapter 3.

In general, the constants and their standard errors are similar for the aggregate, disaggregate, and HLM analyses, but there are important differences among the three methods. Overall, the explained variances in the aggregate analysis are extremely large. At the same time, the regression coefficients produced by the aggregate analysis are not reliable and probably not valid. As explained in Chapter 2, the aggregation of data not only decreases the variance in the data, but more importantly, it introduces unpredictable confounding factors into analysis. Thus, the aggregate analysis is extremely vulnerable to the quality of the data including the shape of distributions, the measurement errors, the existence of confounding variables, the number of groups, and the sizes of groups (for example, how many schools and how many students are clustered in each school), the unbalanced school sizes, and the heterogeneity of samples, all of which challenge the reliability and validity of the aggregate analysis. Eventually, an ecological fallacy is committed when the results from the aggregate analysis are used to explain differences among students.

The disaggregate analysis produced more reliable and valid results than the aggregate analysis. The R^2 is much smaller and more reasonable. The coefficients are more reliable than in the aggregate analysis according to the literature and the standard errors of the student-level variables are smaller. However, disaggregate analysis has limitations due to the violation of homogeneity, and unbalanced group sizes, thus likely underestimate the standard errors of the coefficients for the school-level variables. In the HLM analysis, on the other hand, the percentage of variance explained by the model and the coefficient estimates are similar to those in the disaggregate analysis, and even more accurate because the unbalanced school sizes are controlled for by an HLM analysis. Additionally, the standard errors of both the school-level and the student-level variables are more accurate than they are in either the aggregate or the disaggregate analysis. Specifically, the HLM analysis decomposes the total variance among the students into the group-level and the individual-level variance components. In summary, HLM has the highest reliability and validity of the estimated parameters when dealing with clustered data. Because of these advantages, the HLM analysis provides the best estimates of the effects of the independent variables on the mathematical achievement of the students in the Canadian TIMSS 2007 data.

The HLM Analysis in More Detail

In last section the HLM analysis is shown to result in more reliable and valid effect parameters than the aggregate analysis when dealing with nested data, and that

HLM analysis is even better than the disaggregate analysis. In this section, the results from the HLM analysis of the Canadian TIMSS 2007 data for British Columbia and Ontario are examined in greater detail. As noted, the independent variables were divided into three groups, the school variables, the teacher/classroom variables, and the student variables. To explicitly assess the effects of each group of variables independently and with the intervention effects from the other sets of variables, eight models with various combinations of the three groups of variables were assessed for British Columbia and Ontario, respectively.

Particularly, Model 1 was a fully unconditional model without any predictors, which was used to test the between-school variation in the students' achievement in mathematics in each province. Thus, this model provided the baseline estimates of the intercepts (or grand means), between-school variances (τ_{00}) (or the variance of mean school scores from the grand mean), and the estimates of the within-school variance (σ^2) (or the deviation of the mean individual scores from the school means), which were all necessary information for answering the research hypotheses specified in Chapter 3. Also, τ_{00} and σ^2 served as points of comparison for the subsequent analyses in the other seven models.

In Model 2 through Model 8, all the school and teacher/classroom variables were centered at their provincial means. As well, the student variables were centered at their provincial means except for the students' gender. The standardized coefficients were calculated by standardizing the variables to their means in each province, a method suggested by Bring (1994). The grand means (intercepts) and their standard errors, the unstandardized regression coefficients and their standard errors, the standardized regression coefficients, and the variance components at each level of the eight HLM analyses for British Columbia are reported in Table 51, and the results of the same analyses for Ontario are reported in Table 52. As suggested in the TIMSS manual (Foy & Olson, 2009), the results reported in these two tables are actually the averages of five estimates, one for each of the five plausible values in mathematics. Additionally, the student samples have been weighed by the *final student weight* provided in the TIMSS data set so that the samples represent all the public school students in both British Columbia and Ontario. Generally, the student's weight reflects the selection probability for student *j* attending school *i*, which is the product of the selection probability of school *i* multiplied by the selection probability of student *j* within school *i*. The *final student weight* is the selection probability for the students after adjusting for those who did not complete the TIMSS test.

Mean Achievement

In Model 1, the fully unconditional model without any independent variables, it was estimated that the provincial average mathematics score was 514.40 in British Columbia and 527.14 in Ontario. Obviously, students in Ontario had significantly higher average achievement scores, about 13 points higher, than students in British Columbia. Additionally, the 95% confidence interval of school means in Ontario (520.61–533.67) was slightly narrower and barely overlapped with the confidence interval in British Columbia (506.03–522.77), which indicates that schools in Ontario were more homogeneous in their students' achievement than schools in British Columbia.

Table 51

Results of the HLM analysis in British Columbia

	Model 1		Model 2			Model 3			Model 4		
	Unstandardized		Unstandardized			Unstandardized			Unstandardized		
	B	SE	B	SE	Beta	B	SE	Beta	B	SE	Beta
Grand Means	514.40	4.27	516.89	4.47		514.20	4.15		511.77	3.34	
School Variables											
Physical Resources						-1.24	2.34	-.04			
Instruc. Resources						.97	.96	.06			
Average SES						8.87	5.83	.11			
Administration						.43*	.17	.11			
Teacher/Classroom Variables											
Gender									.39	8.02	.00
Experience									3.49	1.91	.09
Number and algebra									2.85***	.63	.20
Geometry									-.05	.73	.00
Data and chance									-.49	1.17	-.02
Assign. of homework									5.94	3.76	.07
Attitudes									-1.34	1.51	-.03
Homo. of students									6.44***	1.22	.30
Physical resources									.74	2.94	.01
Instruct. Resources									-1.18	1.43	-.04
Student Variables											
Gender			-5.10	3.37	-.04						
Mother education			.69	.91	.02						
Number of books			9.85***	1.14	.17						
Instru. motivation			3.38***	.63	.11						
Edu. expectations			6.65***	.99	.16						
Effort in homework			-11.44***	1.71	-.14						
		SD		SD			SD			SD	
School-Level Variance τ_{00}	1723.61	41.52	1370.63	37.02		1662.97	40.78		1080.40	32.87	
Student-Level Variance σ^2	3414.22	58.43	2974.87	54.54		3413.17	58.42		3414.35	58.43	

$*p \leq .05$, $***p \leq .001$.

Table 51 cont.	Model 5			Model 6			Model 7			Model 8		
	Unstandardized			Unstandardized			Unstandardized			Unstandardized		
	B	SE	Beta	B	SE	Beta	B	SE	Beta	B	SE	Beta
Grand Means	516.81	4.42		514.67	3.81		511.60	3.24		514.62	3.75	
School Variables												
Physical Resources	-1.38	2.08	-.04				-1.12	1.58	-.03	-1.39	1.47	-.04
Instruc. Resources	.70	.82	.04				.35	1.05	.02	.08	.95	.00
Average SES	5.97	5.60	.07				6.95	4.75	.09	4.80	4.80	.06
Administration	.32*	.17	.08				.41*	.16	.11	.31*	.16	.08
Teacher/Classroom Variables												
Gender				1.44	7.37	.00	3.84	8.03	.03	4.10	7.64	.03
Experience				2.91	1.82	.07	3.66*	1.82	.09	3.04	1.72	.08
Number and algebra				2.45***	.56	.17	2.88***	.61	.20	2.39***	.56	.17
Geometry				.08	.64	.00	.12	.75	.01	.24	.65	.02
Data and chance				-.02	1.02	.00	-.37	1.10	-.02	.09	.96	.00
Assign. of homework				5.00	3.37	.06	7.15	3.95	.09	5.82	3.67	.07
Attitudes				-1.72	1.41	-.04	-3.15	1.89	-.08	-2.99	1.70	-.07
Homo. of students				5.57***	1.13	.26	5.96***	1.29	.28	5.32***	1.25	.25
Physical resources				-.79	2.73	-.02	1.37	2.93	.03	-.33	2.84	-.01
Instruct. Resources				-.57	1.33	-.02	-1.20	1.49	-.04	-.43	1.46	-.01
Student Variables												
Gender	-5.22	3.37	-.04	-5.12	3.37	-.04				-5.22	3.36	-.04
Mother education	.58	.91	.01	.50	.91	.01				.41	.90	.01
Number of books	9.74***	1.14	.17	9.73***	1.13	.17				9.65***	1.15	.17
Instru. motivation	3.42***	.64	.11	3.37***	.63	.11				3.42***	.63	.11
Edu. expectations	6.61***	.99	.16	6.54***	.97	.15				6.52***	.98	.15
Effort in homework	-11.43***	1.72	-.14	-11.47***	1.69	-.14				-11.43***	1.69	-.14
		SD			SD			SD			SD	
School-Level Variance τ_{00}	1359.62	36.87		910.16	30.17		1053.96	32.46		902.64	30.04	
Student-Level Variance σ^2	2973.95	54.53		2973.53	54.53		3413.50	58.43		2972.98	54.53	

$*p \leq .05$, $***p \leq .001$.

Table 52

Results of the HLM analysis in Ontario

	Model 1		Model 2			Model 3			Model 4		
	Unstandardized		Unstandardized			Unstandardized			Unstandardized		
	B	SE	B	SE	Beta	B	SE	Beta	B	SE	Beta
Grand Means	527.14	3.33	531.18	3.23		527.18	2.87		527.88	3.20	
School Variables											
Physical Resources						-5.36***	1.49	-.18			
Instruc. Resources						2.30**	.80	.14			
Average SES						12.37***	3.40	.16			
Administration						.06	.18	.02			
Teacher/Classroom Variables											
Gender									8.30	6.20	.06
Experience									1.41	2.51	.04
Number and algebra									.11	.62	.01
Geometry									1.05	.66	.07
Data and chance									-.80	1.63	-.03
Assign. of homework									3.27	3.36	.04
Attitudes									4.91**	1.61	.15
Homo. of students									1.26	1.19	.06
Physical resources									-1.66	1.91	-.04
Instruct. Resources									-1.16	1.61	-.05
Student Variables											
Gender			-9.96**	3.64	-.07						
Mother education			2.37	1.23	.06						
Number of books			10.52***	1.7	.19						
Instru. motivation			3.19***	.84	.10						
Edu. expectations			7.69***	1.23	.18						
Effort in homework			-15.65***	2.16	-.20						
		SD		SD			SD			SD	
School-Level Variance τ_{00}	787.74	28.01	497.04	22.29		555.39	23.57		662.16	25.73	
Student-Level Variance σ^2	3696.35	60.80	3170.92	56.31		3690.98	60.75		3694.06	60.78	

**$p \le .01$, ***$p \le .001$.

Table 52 cont.	Model 5			Model 6			Model 7			Model 8		
	Unstandardized			Unstandardized			Unstandardized			Unstandardized		
	B	SE	Beta	B	SE	Beta	B	SE	Beta	B	SE	Beta
Grand Means	530.76	2.96		531.62	3.15		527.60	2.86		526.40	4.21	
School Variables												
Physical Resources	-4.13***	1.23	-.14				-4.76***	1.40	-.16	-3.81**	1.29	-.13
Instruc. Resources	1.22	.66	.07				2.09**	.78	.12	1.22	.66	.07
Average SES	10.42***	2.86	.13				11.87***	3.45	.15	10.45***	3.02	.13
Administration	.02	.15	.00				-.08	.19	-.02	-.04	.17	-.01
Teacher/Classroom Variables												
Gender				8.04	5.29	.03	8.11	5.48	.06	7.67	4.83	.05
Experience				.80	1.88	.06	2.08	2.00	.05	1.15	1.67	.03
Number and algebra				.12	.52	.02	-.09	.57	-.01	-.06	.49	.00
Geometry				.82	.53	.01	.91	.62	.06	.73	.51	.05
Data and chance				-1.23	1.38	.06	.46	1.49	.02	-.08	1.26	.00
Assign. of homework				3.84	2.89	.05	5.64*	2.65	.08	5.53*	2.46	.07
Attitudes				3.12*	1.27	.10	3.30*	1.43	.10	1.91	1.20	.06
Homo. of students				.87	.89	.04	.22	.86	.01	-.01	.68	.00
Physical resources				-2.79	1.63	-.05	-2.14	1.69	-.05	-3.09*	1.54	-.07
Instruct. Resources				-.50	1.23	-.07	-.74	1.29	-.03	-.12	1.05	-.01
Student Variables												
Gender	-9.80**	3.56	-.07	-9.75**	3.61	-.07				-9.46**	3.57	-.07
Mother education	2.14	1.24	.05	2.24	1.24	.06				2.00	1.24	.05
Number of books	10.10***	1.69	.19	10.46***	1.69	.19				10.16***	1.68	.19
Instru. motivation	3.20***	.84	.10	3.115***	.85	.10				3.07***	.85	.10
Edu. expectations	7.72***	1.22	.18	7.53***	1.23	.18				7.63***	1.24	.18
Effort in homework	-15.52***	2.13	-.20	-15.82***	2.09	-.20				-15.72***	2.10	-.20
		SD			SD			SD			SD	
School-Level Variance τ_{00}	367.81	19.18		438.48	20.94		476.08	21.82		322.94	17.97	
Student-Level Variance σ^2	3168.24	56.29		3170.80	56.31		3691.53	60.76		3168.20	56.29	

$*p \leq .05$, $**p \leq .01$, $***p \leq .001$.

Overall, when adding the independent variables in Model 2 to Model 8, the mean estimates were relatively stable. Since the students' gender was not centered in the analysis, it was estimated that the average achievement score of the males in mathematics was 516.89 in British Columbia and 531.18 in Ontario (Model 2). At the same time, the average achievement score of the females was 511.79 in British Columbia and 521.22 in Ontario, slightly lower than the average scores of the males. In British Columbia, adding the teacher/classroom variables has decreased the grand mean slightly to 511.77 (Model 3). However, in Ontario, adding the school variables and the teacher/classroom variables did not change the mean estimates. Similarly, the mean estimates in Model 5 through Model 8 with various combinations of the three groups of variables were stable compared with Model 2 through Model 4 where each group of variables was separately assessed, which indicates that the three groups of variables are relatively independent of each other.

Variance Components

In Model 1, there was statistically significant variation among the schools in both British Columbia (χ^2=1289.33, df=133, p<.001) and Ontario (χ^2=621.51, df=167, p<.001). In particular, 33.55% [$\tau_{00}/(\tau_{00}+\sigma^2)$] of the total variance in the students' achievement in mathematics in British Columbia and 17.51% in Ontario was between schools. Evidently, much larger variance of the students' achievement in mathematics was observed among schools in British Columbia than in Ontario. Even after all the school, teacher/classroom, and student variables had been included in Model 8, significant between-school variance still existed in both British Columbia (χ^2=783.30, df=119, p<.001) and Ontario (χ^2=346.08, df=153, p<.001). In this model, the between-school variance accounted for

23.29% of the total variance in the students' achievement in British Columbia, but only 9.25% of the total variance in the students' achievement in Ontario. Nevertheless, the independent variables have explained 24.57% of the total variance in the students' achievement in mathematics in British Columbia and 22.14% of the total variance in the students' achievement in Ontario.

In Model 2 through Model 5, it is noted that the school, teacher/classroom, and student variables have acted differently in British Columbia than in Ontario. More specifically, the six student variables have explained similar amounts of variance in the students' achievement in the two provinces though slightly less in British Columbia (14.52%) than in Ontario (18.20%). At the same time, they have explained similar amounts of the within-school variance, 12.87% in British Columbia and 14.21% in Ontario, but different amounts of the between-school variances, 20.48% in British Columbia and 33.67% in Ontario. Obviously, more between-school variance in Ontario has been explained by the students' individual characteristics, which demonstrates that different average achievement among schools is a reflection of the heterogeneity of students among those schools. In other words, the students in Ontario are less randomly distributed among schools than the students in British Columbia.

In Model 7, it is noted that a similar amount of between-school variance has been explained by the school-level variables, 38.85% in British Columbia and 39.56% in Ontario. However, it is surprising that the schools and teachers/classrooms variables were playing different roles in the two provinces. Specifically, in Model 3 the school-level variables barely explained any between-school variance in British Columbia (3.51%) while they explained a large amount of between-school variance in Ontario (29.23%).

However, in Model 4, the teacher/classroom variables explained a large amount of the between-school variance in British Columbia (37.32%) and much less in Ontario (15.62%). Combining the school, teacher/classroom, and the student variables in Model 5 through Model 8 did not affect the amount of variance explained by the three sets of variables in comparison with Model 2 through Model 4, which confirms that the three sets of variables are relatively independent in both provinces.

School Variables

The school variables affected the students' achievement in different ways in British Columbia and Ontario. The effects of the school variables were essentially stable across Model 3, Model 5, Model 7, and Model 8, which indicates that they are independent of the student and the teacher/classroom variables in both provinces. However, there were some small mediating effects by both the student and the teacher/classroom variables. The largest mediating effect by the student variables was for the instructional resources of the schools in Ontario, which decreased from being statistically significant and moderate in Model 3 (Beta=.14, $p<.01$) to about half that size and not significant in Model 5 (Beta=.07, $p>.05$).

After all the teacher/classroom and student variables had been controlled for in Model 8, it is shown that in British Columbia, only the school administration variable had a statistically significant and moderate effect on the achievement of the students (Beta=.08, $p<.05$), which is consistent with the research literature (Andrews & Soder, 1987; Bossert, Dwyer, Rowan, & Lee, 1982; Hallinger, 2003; Hallinger & Heck, 1996). Of course, this significant effect may have resulted because the school administration variable had a larger variance in British Columbia than in Ontario. It is surprising, however, that the average SES of the schools did not have a statistically significant effect in British Columbia, which is contradictory to the fact that in Model 2 the composition of the students' characteristics explained 20.48% of the between-school variance. Theoretically, the educational environment created by students in schools seems to play an important role in their achievement. Amongst the commonly used measures, the approximate percentage of the students coming from economically disadvantaged homes, is probably not a very good scale, especially in British Columbia where the school size tend to be large. In Ontario, on the other hand, two school variables had statistically significant and moderate effects on the students' achievement. Specially, the higher average SES of the schools judged by the principals had a statistically significant and moderate effect on the students' achievement (Beta=.13, p<.001). While this finding is consistent with the research literature, it is surprising that the physical resources of the schools had a negative, but statistically significant, effect on the students' achievement (Beta= - .13, p<.001).

Teacher/Classroom Variables

Like the school variables, the teacher/classroom variables also affected the students' achievement differently in British Columbia and Ontario. The teacher/classroom variables were relatively independent of the students' characteristics except for a few small mediating effects resulting from those characteristics. In Ontario, the effect of the teachers' attitudes has been mediated by both the student and the school variables indicated by the decreased standardized coefficients from Model 4 (Beta=.15, p<.01) to Model 8 (Beta=.06, p>.05). Additionally, the effects of the teachers' assigning homework and the physical resources in the classrooms have been slightly suppressed by the student and the school variables.

After all the teacher/classroom and student variables have been controlled for in Model 8, in British Columbia, the homogeneity of students had a statistically significant and strong effect on their achievement (Beta=.25, p<.001), which is a little surprising because the schools in British Columbia had more homogeneous students than in Ontario. From Model 2, it is shown that 20.48% of the between-school variance in British Columbia and 33.67% of the between-school variance in Ontario has been explained by students' characteristics, which indicates a higher heterogeneity among schools in Ontario than in British Columbia in the students' backgrounds including their gender, SES, educational expectations, instrumental motivation, and their effort in completing the homework. However, the heterogeneity of students did not have an impact in Ontario. Additionally, the instructional time spent on teaching number and algebra also had a statistically significant and strong effect (Beta=.17, p<.001), which is probably caused by the fact that around 70% of the test items in the math exam focused on number and algebra but not geometry and data and chance. In Ontario, two variables, the assignment of homework from the teachers (Beta=.07, p<.05) and the physical resources in the classrooms (Beta= -.07, p<.05), had small but statistically significant effects on the students' achievement, which were both suppressed by the student and school variables.

Student Variables

Unlike the school and the teacher/classroom variables, the effect of student variables was consistent in both provinces, and they were largely independent of the school and the teacher/classroom variables. This demonstrates, again, the importance of the students' characteristics in their academic achievement. Furthermore, the similarities between the two provinces show that a HLM analysis is appropriate for these nested data sets. More specifically, in Model 8, the gender difference was only statistically significant in Ontario (Beta= -.07, $p<.01$), and not in British Columbia (Beta= -.04, $p>.05$). Nevertheless, the effect in both provinces was negative which shows that males do slightly better in mathematics than females. Meanwhile, an increased number of books in the students' homes significantly increased their achievement in both British Columbia (Beta=.17, $p<.001$) and Ontario (Beta=.19, $p<.001$). In addition, students with higher instrumental motivation significantly outperformed students with less motivation in both British Columbia (Beta=.11, $p<.001$) and Ontario (Beta=.10, $p<.001$). Likewise, students with higher educational expectations significantly outperformed students with lower expectations in both British Columbia (Beta=.15, $p<.001$) and Ontario (Beta=.18, $p<.001$). Finally, the time students spent on completing homework was significantly associated with lower achievement for students in both British Columbia (Beta= -.14, $p<.001$) and Ontario (Beta= -.20, $p<.001$), which suggests that students who were not doing well generally spent more time on their homework. All these findings are supported by the research literature (see Chapter 3).

Answering the Hypotheses

This chapter has compared the difference between aggregate analysis, disaggregate analysis, and HLM analysis by using a large-scale data set. Further, the HLM method has been used to analyze the Canadian 2007 TIMSS data to assess school effectiveness when a number of important student variables have been controlled. By doing so, this chapter has addressed the four hypotheses proposed in Chapter 3.

Hypothesis 1. The aggregate analysis gives the most biased estimates of school effectiveness on the students' academic achievement.

As explained in Chapter 2, the aggregation of data not only decreases the variance in the data, but more importantly, it introduces unpredictable confounding factors into the analysis. Thus, aggregate analysis is extremely vulnerable to the distributions in the data. Consequently, the parameter estimates in aggregate analysis are quite different from the same estimates in the disaggregate and the HLM analyses, which is probably more evident for the student variables since their parameter estimates have been relatively consistent in the literature. Generally, the regression coefficients produced by aggregate analysis are not reliable and probably not valid. At the same time, because of the decreased variance in the variables, the R^2 in the aggregate analysis is extremely large. Generally, the smaller the number of aggregate units, the larger the R^2 becomes. The aggregate analysis used in many studies may contribute to the mixed results about the effects of the physical resources in schools in the research literature.

Hypothesis 2. The HLM gives the most accurate estimates of school effectiveness on the students' academic achievement.

HLM analysis decomposes the total variance among the students into group-level and individual-level components and thus assesses the schools' impact while controlling for individual differences. This gives HLM analysis several advantages. In HLM analysis, the parameter estimates are similar to those in disaggregate analysis but more accurate because the heterogeneity of the samples and unbalanced school sizes are considered. Additionally, the standard errors of both the school-level and the student-level variables are more accurate than in either aggregate or disaggregate analysis. Overall, HLM analysis has the highest reliability and validity of the estimated parameters when dealing with nested data, such as the TIMSS data.

Hypothesis 3. There are significant provincial differences between British Columbia and Ontario in the students' academic achievement.

Students in Ontario have significantly higher scores than students in British Columbia even after more of the lower performed students and their schools were excluded from this study due to the school-level missing values. This finding indicates that, overall, schools in Ontario are more effective than schools in British Columbia. Interestingly, a similar amount of between-school variance is explained by the school and teacher/classroom variables in the two provinces. However, in British Columbia, the differences among schools are small while the classrooms within schools vary considerably, largely because of the differences in the instructional time the teachers spend on teaching number and algebra and with the homogeneity of students. On the contrary, in Ontario, the differences among schools in physical resources and the average SES of their students are important, while classrooms within schools are similar.

Hypothesis 4. The physical resources of schools, which are evaluated by school principals and classroom teachers, have a positive impact on the students' academic achievement in both British Columbia and Ontario.

The analysis shows, however, that the physical resources evaluated by the school principals have no impact on the achievement of students in British Columbia and only a moderate negative impact in Ontario. Additionally, the physical resources evaluated by the classroom teachers have no impact on the students' academic achievement in either British Columbia or Ontario. This finding can be explained in several ways. First, the variance of this variable may be small in all Canadian provinces. Second, the adequacy of the physical resources as evaluated by the school principals and teachers may not be accurate especially as they relate to the performance of the students. Finally, the physical resources may be above the adequacy level, meaning that they would affect the students' academic achievement only when they are below that level. Overall, physical resources have no impact on the students' learning as evaluated by the TIMSS test scores in mathematics.

Summary

After assessing two basic assumptions about the Canadian 2007 TIMSS data, the missing values and collinearity, and comparing aggregate, disaggregate, and HLM analyses, the third section in this chapter presents the results of the HLM analysis of the Canadian TIMSS 2007 data for British Columbia and Ontario in much greater detail. The analysis demonstrates that the students' individual characteristics have the most important effect on their achievement in mathematics and these effects are highly consistent regardless of the classroom, school, or province the students are from. Compared to the students' characteristics, the school and the teacher/classroom variables have smaller effects. However, there are moderately large and significant effects from some of the institutional variables. While these variables are not affected by the students' characteristics, they vary considerably by province. Finally, it is confirmed that British Columbia and Ontario differ in the students' average achievement in mathematics even after their individual characteristics have been controlled, which can be partially explained by some of the institutional variables. Based on these findings, the last section in this chapter provides answers to the four hypotheses proposed in Chapter 3. Many of the findings in this section, especially the effect of students' characteristics, are consistent with the literature. Some other findings deserve further discussion, which is presented in the next chapter.

CHAPTER 5

CONCLUSION

This study used multilevel modeling procedures, specifically, Hierarchical Linear Modeling (HLM) and the Canadian TIMSS 2007 data, to evaluate the effect of school variables on the students' academic achievement in mathematics when a number of student variables were controlled. At the same time, the provincial differences between British Columbia and Ontario have been evaluated by analyzing the data in each province separately. In this chapter, I summarize the substantive findings of the dissertation and discuss them in relation to the research literature. I suggest that, currently, the contradictory findings in the research literature about the effect on the students' achievement from a number of school variables, specially the physical resources in schools, are largely due to the analytical procedures that have been used by a number of researchers. Using a more appropriate method, HLM for example, and a large data set, TIMSS for example, this dissertation helps to clarify some of the significant misunderstandings in the current research literature. Even though the physical resources in schools were found to have little, if any, effect on the students' academic achievement in this study, it does not mean that physical resources are not important. Rather, it suggests that the measurement of the physical resources should be improved, specifically from measuring what resources are available to measuring how the resources are used. This change in measurement would help us understand better what is happening in schools. Before discussing findings I first review the content of each chapter.

A Review of the Study

In Chapter 1, I introduce the problem examined in this dissertation, the effect of school variables on students' academic achievement when a number of student variables were controlled using HLM and the Canadian TIMSS 2007 data set. With improved modeling procedures, this dissertation helps clarify some of the contradictory findings in the current literature on school effects, which may be caused by inappropriate analytical procedures. I also listed four limitations of this study.

In Chapter 2, I explain why and how researchers can study cross-level effects of schooling using multilevel modeling techniques. It is essential to use multilevel techniques in school effectiveness research because of the multilevel nature of schooling and the methodological problems evident in the current research. I explain that multilevel analysis is more appropriate than other methods for analyzing these nested educational data. The second section explains that how multilevel modeling can be used to study school effectiveness, making multilevel analysis, specifically HLM, appropriate for the analysis undertaking in this project.

In Chapter 3, the methodology chapter, I provide a description of the large samples of schools, mathematics teachers, and eighth grade students that were included in this study. In addition, I present the dependent and independent variables included in the research, their theoretical significance, the items used to measure the variables, and their descriptive statistics. Further, four hypotheses are presented that were proposed based on the review of literature and the descriptive statistics.

In Chapter 4, I first review the analytical methods that were used to examine missing values in some variables and collinearity among the independent variables. I then report the results produced by the aggregate analysis, disaggregate analysis, and HLM analysis. The evidence showed that HLM provided more accurate estimates than either the aggregate or the disaggregate analysis. Consequently, I used two-level HLM analysis to evaluate the effects of school, teacher, and classroom variables on achievement scores in mathematics at a higher level while controlling for students' characteristics at a lower level. Most importantly, this chapter provides answers to the four hypotheses tested in this study. First, aggregate analysis gives the most biased results about the effectiveness of schools on students' academic achievement, including biased results for the effect of schools' physical resources. Second, HLM gives the most accurate estimates using the nested TIMSS data. Third, there are significant provincial differences between British Columbia and Ontario in the students' academic achievement. Finally, the physical resources of the schools, which have been evaluated by school principals and classroom teachers, have no positive impact on the students' academic achievement in either British Columbia or Ontario.

In Chapter 5, I first review the main arguments of each chapter. The following discusses the limitations of this study posed by missing values in the data. Given the constraints of excluding some schools and students, I make the case that the findings of this study still could be generalized to other Canadian provinces. Following this, I discuss some of the findings in relation to the research literature; specifically, I examine the biases caused by aggregation of data, some issues in school effectiveness research, and the provincial differences. Finally, I suggest some further research that could be conducted to answer questions raised in this study.

Effects of Missing Data

Missing data is a common problem in quantitative research studies (Peugh & Enders, 2004). Rubin (1976) defined missing data as “missing completely at random” (MCAR) where the missing values on a particular variable are unrelated to other variables in the data set. Essentially, the observed data set represents a random sample of the hypothetically complete data set. Rubin’s “missing at random” (MAR) means that the missing values on a variable can be related to other variables, but still unrelated to the underlying values of this variable. Of course, MAR is an assumption and may be difficult to verify. Unfortunately, in survey research, missing data are often not random. A “missing not at random” (MNAR) problem results when the probability of missing values on a variable is related to the underlying values of this variable. In educational research, the refusal or inability to respond to questions may be correlated with the characteristics of the respondents, such as the schools’ administration, geographic location, or average achievement of students as well as their interest or their SES, etc. In these cases, excluding the cases or the variables with large amounts of missing values without carefully assessing the missing values may produce biased parameter estimates.

In this dissertation, the MNAR type of missing data was evident, but I carefully assessed the differences between the schools and students that have been excluded and those who remained in the study in order to identify some possible biases. As explained in Chapter 3, there were missing values in the Canadian TIMSS 2007 data, especially for the school data in British Columbia and the student data in both British Columbia and Ontario. In fact, the principals' responses to the questionnaire about their school and teachers from 16 schools (10.67% out of 150 schools) in British Columbia with 396 students (9.30% of the students) and seven schools (3.98% out of 176 schools) in Ontario with 113 students (3.28% of the students) were missing. If these schools were random, the effect of the missing values would not be a problem for analysis. To assess if the missing schools were random, I compared the students in the schools with missing values to the other schools. T-tests indicated that in both provinces, students with school-level missing values had significantly lower achievement scores than the other students. Nevertheless, in British Columbia the two groups of students did not differ in their gender, SES, or their educational expectations. In Ontario, on the other hand, students from schools with missing values had significantly lower SES and lower educational expectations than the other students. Because there were only seven schools which were missing school-level values in Ontario, the comparison between schools with and without missing values in this province may not be as reliable as in British Columbia.

Obviously, excluding the 16 schools and their 396 students in British Columbia and seven schools and their 113 students in Ontario from analysis may affect the results differently for the two provinces. More specifically, in British Columbia, about 10% of the schools and students that were not performing well on the mathematics exam have been dropped from the study though these students may not differ from the other students on their individual characteristics. However, because the school-level values are missing, it is impossible to evaluate what school factors affected the students' achievement. Because these excluded schools, the results may only be applicable to the better performing schools in British Columbia. Additionally, the missing values in these schools probably have reduced the variance in the school-level variables, thus reducing the statistical power of the tests. On the contrary, in Ontario, only around 4% of the schools with poorly performing students have been excluded from analysis, which probably has only a minor impact on the findings for this province. It is worth noting that the analysis about the provincial differences in this study was actually comparing all the schools in Ontario with those slightly better performing schools in British Columbia.

For the remaining students in both British Columbia and Ontario, the mothers' and the fathers' education of the students both had about 40% missing values, a problem probably caused by the students' checking "I don't know" to the question about their parents' education. These missing values were examined to determine if they occurred at random. It was found in both provinces that the students who reported not knowing their mothers' education were more likely to report not knowing their fathers' education. However, the associations between the students' response to this item and not answering items about gender and educational expectations were very small. There were also significant differences between the students with and without missing values on their mothers' education. In both provinces, the students who did not report their mothers' education had significantly lower mathematics achievement scores than those who reported their mothers' education. In addition, the student who did not report their mothers' education had significantly lower educational expectations than the other. Because of the positive relationship between the mother's education and student's academic achievement, the parents of those students who have missing values on their parents' education probably received less education. Consequently, the students who did not provide information about their parents are more likely to have lower SES than the students who provided information about their parents. Because the mother's education is theoretically an important indicator of a student's SES and this variable should be controlled in school effectiveness research, those students with missing values on their mothers' education unfortunately were excluded from the analysis. Consequently, removing around 40% of the lower performing students not only has significantly reduced the sample sizes in both British Columbia and Ontario, but it has also reduced the variance in some variables included in analysis. Consequently, this may restrict the generalizability of the findings in this dissertation.

Nevertheless, excluding schools and students with large amounts of missing values can reduce the errors in further analyses by decreasing the missing values of each variable to a manageable amount. Consequently, the estimated results would be more accurate and reliable. After excluding the schools and students that had missing values, 134 schools with 2,392 students in British Columbia and 169 schools with 2,080 students in Ontario were included in the study. The sample sizes were adequate for the aggregate, disaggregate, and especially for the HLM analysis which needs a large sample of both schools and students. The findings in this study are consistent with the best research literature. They have helped clarify some of the mixed results in this literature. Thus, the results are mostly acceptable, but some caution is needed when generalizing to all students and schools in these two provinces.

Effects of the Method of Analysis

In the Canadian educational system, students are grouped within classrooms, classrooms are grouped within schools, schools are grouped within districts, and districts are grouped within provinces. Briefly, each higher level provides the lower levels with contextual conditions that potentially affect students' academic achievement. Thus, school effectiveness is a cross-level effect because the educational influences on students occur in the groups to which the students belong. Through both direct and indirect cross-level effects, students' academic achievement is affected.

This multilevel view of schooling implies that school characteristics cannot be conceived of as being treatments that are uniformly administered to all students within a specific school, nor can they be conceived of as being a constant to each student in a specific school. On one hand, the resources for instruction and the opportunities for learning vary across schools, classrooms, and students. On the other hand, given the same opportunities for learning, students often respond quite differently. As such, an adequate conceptualization of the effects of schooling must include not only how schools differentially allocate resources and create learning environments for students, but also how students differentially respond to those environments.

Often this cross-level perspective is modeled in two ways, using disaggregate and aggregate analyses, both of which evaluate the cross-level impact of variables at a single level. Unfortunately, both procedures are flawed as demonstrated in this dissertation using a large data set from two Canadian provinces, British Columbia and Ontario. The disaggregate analysis ignores the violations of important assumptions, such as the normal distribution of data, the homogeneity of variance, and the independence of the observations. As a result, the standard errors calculated are misleadingly small, and consequently the null hypothesis is rejected more often than expected, especially for the institutional variables. In other words, the institutional variables are more likely to show statistically significant effects on the students' academic achievement in the disaggregate analysis than they actually have if they could be measured accurately.

A more significant problem, however, exists when the data are aggregated, which has both methodological and theoretical difficulties. Normally, the ecological coefficients from aggregated data are much higher than the coefficients estimated at the individual level. The more the data are aggregated due to smaller number of groups, the larger the estimated coefficients (Robinson, 1950). Most importantly, the coefficients are often not realistic because they cannot link the individual outcome to the individual's exposure to the contextual variables, a problem that makes it almost impossible to infer relationships at the individual level. Theoretically, modeling using aggregate analysis assumes that, within a multilevel hierarchy, each student in a classroom or in a school has been equally exposed to the learning opportunities and each responds similarly to those opportunities. As explained in Chapter 2 and demonstrated in Chapter 4, this assumption is unrealistic because individuals vary substantially given the same learning opportunities. Both methodological and theoretical reasons imply that the pattern of correlations at the higher, or aggregated, level cannot be simply attributed to individuals. Simply equating the two commits an ecological fallacy.

Unfortunately, many studies using aggregate analysis seem to disregard ecological fallacy in understanding the disparity of different groups of students in schools with different physical resources. These studies attribute the effect of the physical resources of schools upon students by connecting the schools' physical resources with aggregated learning outcomes and thus, by disregarding the fact that the level the schools' physical resources are measured at is different from the level of the students. However, multilevel analysis, asks if the observed correlation between the physical resources of schools and the learning outcome of students can be explained by the intervention of other variables or if there is a truly causal relationship between the two. When this type of analysis is conducted, the association between the schools' physical resources and the students' learning outcome is not as high as in the aggregate analysis. But, in aggregate analysis, questions remain about the extent to which statistical artifacts have been mistaken for real effects. There may be a very simple explanation for some, or all, of the reported associations between the schools' physical resources and the students' learning outcome. A positive relationship between the two can arise at the aggregate level even if the schools' physical resources have no effect on the students' academic achievement.

First, the relationship between the schools' physical resources and the students' achievement is not likely to be linear. On the contrary, the function is expected to be curvilinear so there is a diminishing return in achievement to the quality of the physical resources in schools. In other words, resources matter when there are few resources and probably less when there are lots of resources. In the aggregated data, because of this nonlinear correspondence between the schools' physical resources and the students' achievement, even if the increased physical resources of schools have no impact upon the students' achievement at the higher end of a distribution, schools with more physical resources will, when all other things are equal, appear to have higher average student achievement.

Second, it is questionable whether or not the correlation between the physical resources of schools and students' academic achievement is causal. There are three possible interpretations of a correlation between the physical resources of schools and students' academic achievement: 1) more physical resources in schools increase students' achievement, 2) higher achievements of students improves the physical resources of schools, or 3) they are independent of each other but both are related to a third variable. An aggregate analysis can hardly rule out any of these three interpretations because the confounding variables linking the two variables are not included in the analysis. Schools with above average physical resources are often found in newer and wealthier communities. Thus, most students in these schools are from relatively affluent families. They are likely to have well educated parents with good occupations and incomes and so the children are likely to have higher academic achievement in schools. As a result, those schools are likely to have, on average, higher student achievement. As well, schools with more students from middle class families who tend to behave well will suffer less damage to their schools' facilities. As a result, fewer resources are required to maintain and improve the physical conditions of the schools. Overall, the relationship between the schools' physical resources and the students' academic achievement in an aggregate analysis cannot prove a causal relationship without controlling for a number of important confounding variables. Some researchers claim that the teachers' better attitudes, the students' better behavior, and their increased satisfaction are the result of improved physical resources of schools and are important mediating variables that increase the students' success (Earthman & Lemasters, 2009; Uline & Tschannen-Moran, 2008; Uline, Tschannen-Moran, & Wolsey, 2009). Unfortunately, the assumed impacts of these variables on students' learning have not been empirically demonstrated by using an appropriate methodology. In summary, the relationship observed at an aggregate level may misrepresent the true relationship between the physical conditions of schools and the academic achievement of students. Consequently, in aggregate analysis, an ecological fallacy may be evident and undermine the conclusions of a potential causal relationship between variables in a number of studies (Berner, 1993; Buckley, Schneider, & Shang, 2004; Cash, 1993; Earthaman, 2004; Lewis, 2000; Roberts, 2009; Tanner, 2000; Tanner & Lackney, 2006; Uline & Tschannen-Moran, 2008; Uline, Tschannen-Moran, & Wolsey, 2009).

Nevertheless, recently developed multilevel methods are directed at properly assessing the cross-level interactions between individual variables and group variables by analyzing the effects at their proper levels. These methods, such as HLM, properly conceptualize the multilevel view of schooling. That is, students are influenced by classrooms managed by teachers within schools, districts, and provinces. And these factors, at different levels, affect students differentially. Consequently, the results of such analysis are more reliable and valid, and thus the interpretations of their findings are more accurate. Additionally, they are more interpretable than either disaggregate or aggregate analysis of the same data set because the multilevel analysis allows researchers to calculate the proportion of the variances in the dependent variables that are attributable to both the lower (student) and higher (school) levels. Using this research method, school effectiveness is thus decomposed into individual student variance, teacher and classroom variance, school variance, district variance, and provincial variance. The importance of variables across levels can then be compared. Thus, although this study has produced more reliable and realistic estimates especially for the school variables, we need to be cautious in interpreting its findings.

Student, Teacher/Classroom, and School Effects

In this analysis, students' individual characteristics were found to have the most important effects on academic achievement in mathematics, which is consistent with the research literature (Bosker & Witziers, 1996; Bryk & Raudenbush, 1992; Coleman, Campbell, Hobson, McPartland, Mood, Weinfeld, & York, 1966; Creemers, 1994; Rowe & Hill, 1998; Luyten, 1994; Madaus, Kellaghan, Rakow, & King, 1979; Scheerens & Bosker, 1997; Stringfield & Teddlie, 1989; Townsend, 2007). More specifically, the gender difference in mathematics favoring males is very small. Also, students from higher SES homes, indicated by the number of books in their homes, are likely to be more successful in schools than students from lower SES homes. In addition, students with higher educational expectations and higher motivation tend to have higher achievement scores than other students. Finally, the time spent on completing homework is negatively related to the students' academic achievement, which probably indicates that it takes less time for the higher-achieving students to finish their homework. It may look surprising that the mother's education did not have a statistically significant impact on their children's academic achievement. However, this may be due to the restricted variance in this variable in both British Columbia and Ontario. Unfortunately, students, who did not report their mothers' education and whose mothers may actually have lower education, were dropped from this study. As a result, the remaining students' mothers tend to have similar and slightly higher education.

Compared to students' characteristics, the school and the teacher/classroom variables have relatively small effects, which is also consistent with the research literature (Coleman, et al., 1966, Goldhaber, 2002; Hanushek, Kain, & Rivkin, 1998; Hattie, 2009; Meyer, 2001; Nye, Konstantopoulos, & Hedges 2004; Townsend, 2007; Wayne & Youngs, 2003; Webster, Mendro, Orsak, & Weerasinghe, 1996). There are, however, moderately large and significant effects from some of the institutional variables, and these vary by province. In British Columbia, the homogeneity of students in the classrooms and the instructional time that the teachers spend on teaching number and algebra have statistically significant and moderate to strong effects on the students' academic achievement. On the contrary, in Ontario, the average SES of students in school has a moderate effect on their achievement. The SES variable does not have an impact on achievement in British Columbia. However, schools in British Columbia are larger, and it may be difficult for principals to answer the item in the questionnaire about the approximate percentage of students coming from economically disadvantaged homes. Perhaps the reliability and validity of the conventional measurement of school SES are rather low in British Columbia. Actually some researchers have claimed that using the average SES of students in school to measure the school SES is a compositional effect or a statistical artifact rather than a contextual effect (Alexander, et al., 1979; Hauser, 1970, 1971; Nash, 2003). Another interesting finding is that the physical resources in schools did not have an impact in British Columbia but they had a moderate negative impact in Ontario. This finding is different from the research literature that shows a positive effect of physical resources (Berner, 1993; Buckley, Schneider, & Shang, 2004; Cash, 1993; Earthaman, 2004; Lewis, 2000; Roberts, 2009; Tanner, 2000; Tanner & Lackney, 2006;

Uline & Tschannen-Moran, 2008; Uline, Tschannen-Moran, & Wolsey, 2009), which has been demonstrated with aggregate analysis. Nevertheless, the results in this study are consistent with other studies that have used multilevel research methods (Bowers & Urick, 2011; Cervini, 2009; Ma & Klinger, 2000; Wei, Clifton, & Roberts, 2011).

Even though the physical resources in schools have little impact on the students' academic achievement, it does not mean that physical resources are unimportant. Rather, physical resources are one factor among many that affect the complex relationships of teaching and learning outcomes. In fact, the research literature premises that the physical resources mediate the relationship of teaching and learning. However, it is arguable that any improvement in the physical conditions of schools impacts more on the comfort and satisfaction of students than on their learning (Lomas, 2005; Nair & Fielding, 2005). In fact, studies using the multilevel methods of the physical resources suggest that an excessive improvement in some specific physical conditions of schools is unlikely to improve the students' achievement if the physical resources in most schools are already adequate or better as they seem to be in British Columbia and Ontario.

Many studies refer to a number of ways that specific physical conditions impact upon the students' learning. Environmental conditions such as noise, temperature, air quality, ventilation, lighting, and cosmetic appearance are often considered when defining the optimal conditions for students to perform well (Berner, 1993; Buckley, Schneider, & Shang, 2004; Cash, 1993; Earthaman, 2004; Lewis, 2000; Roberts, 2009; Tanner, 2000; Tanner & Lackney, 2006; Uline & Tschannen-Moran, 2008; Uline, Tschannen-Moran, & Wolsey, 2009). There are two types of instruments that have been used to measure these conditions. The first one is represented by the Facility Condition

Index (FCI) which is assessed by engineers and architects, and is primarily used to support the management of the schools' assets; it does not relate to the way the physical conditions affect the instructional purpose of schools (Roberts, 2009). The second one is represented by the Commonwealth Assessment of Physical Facilities (CAPE) which is assessed by school administrators and teachers (Cash, 1993); it evaluates the adequacy of the physical conditions of buildings to support their intended instructional purposes (Roberts, 2009). However, these instruments are criticized for their low reliability and validity because of their self-reporting nature (Picus et al., 2005). Nevertheless, the objective measured in the research literature has been the availability of certain physical resources. For example, a typical CAPE assessment questionnaire inquires about the school buildings' age, windows, lighting, grounds, lockers, heating, quality of air, cleaning, painting, repairing, cosmetic appearance, etc. Another typical assessment instrument that is similar to CAPE, the Total Learning Environment Assessment (TLEA) developed by O'Neill (1999), inquires about the adequacy of each component of the physical facilities. What happens if schools already have adequate physical resources for educators and students? The current research design could not answer this question. The mixed findings in the research literature suggest that physical resources are probably reaching a point of diminishing return in provinces where a standard of school buildings has been established and the lighting, air quality, and appearance are likely to be above an adequate level.

In fact, researchers should probably move away from measuring the physical facilities that are available in schools and focus on how those facilities impact on different learning outcomes, or how teachers and students use the facilities pedagogically in ways that improve the students' learning. Probably many intangible factors are more important than those tangible ones that have been measured, especially those that relate to developing an orderly learning environment to support and value the students' academic achievement. In summary, the social practices of teaching and learning that may relate to the use of the physical facilities should be examined more carefully. Unfortunately, this is a large missing piece in the present research on the effects of physical resources in schools. As Heppell and his colleague (2004) say, when we design a space for teaching, we know a lot about minimizing heat loss, but no one knows how to prevent learning loss.

Provincial Effects

There are considerable provincial differences that have been found in this analysis. More specifically, the students in Ontario had significantly higher average achievement scores in mathematics than the students in British Columbia even after more lower performing students were excluded from this study in British Columbia because of school-level missing values. In British Columbia, the differences in achievement of mathematics among schools were smaller than in Ontario while classrooms within schools vary considerably in the amount of time the teachers spent on teaching number and algebra. On the contrary, in Ontario, the differences were among schools and related to physical resources and average SES of students. It is not clear what policies in these two provinces contributed to these differences. Researchers should examine provincial policies to sort out the ones that affect the academic achievement of students.

First, the higher percentages of aboriginal students among the students in British Columbia (9.83%) than in Ontario (3.72%) may contribute to the lower average performance of students in British Columbia. Aboriginal students still perform comparably less well than non-Aboriginal students in Canadian schools (Richards, 2008). Besides, in Western Canada, more students live in small and geographically scattered communities. These students are more likely to have parents who have less education and value the education achievement less. Consequently, their academic achievement in schools is likely to be lower than students in Ontario.

Second, the slightly different organization of the educational system in British Columbia and Ontario may also contribute to the higher performance of the students in Ontario. For example, in the 2009-2010 school year, 89.04% of the K–12 students attended public schools in British Columbia. However, during the same school year, only 67.86% of the students attended public schools in Ontario and most of the others attended Roman Catholic schools. The overall better performance of Roman Catholic schools and the competition between Roman Catholic and public schools may account for some of provincial differences in the students' achievement. However, further research needs to be conducted to determine the specific effect of Roman Catholic versus public schools in Ontario.

Finally, although there are similar large-scale assessments in both British Columbia and Ontario that monitor students' achievement and schools' accountability, the assessment program in Ontario is the responsibility of the Educational Quality and Accountability Office (EQAO), which is at arm's-length from the Ministry of Education. EQAO develops assessment instruments for education programs that are linked to specific learning outcomes in the provincial curriculum. After the examination period, EQAO reports the results to educators, parents, and the public. EQAO seems to organize the testing program better than the ministry of education does in British Columbia. Ontario emphasizes standardized tests more and carefully supervises the tests, which may encourage schools in Ontario to strengthen the teaching of the curriculum. For example, it has been noticed that Ontario has, on average, more school days in a school year than British Columbia. Thus, the students in Ontario achieve higher scores in the TIMSS and PISA than the students in British Columbia.

Even though I found some provincial differences, the results from this study may be generalized to other Canadian provinces because curricula and teaching methods are quite similar across the provinces. In fact, my analysis revealed that in both British Columbia and Ontario the results are generally consistent with the best in school effectiveness research because the students' characteristics were found to be the most important determinants of their academic achievement, while school and teacher effects were also found to be important in providing learning opportunities and positive learning environments. The provincial differences are, in fact, relatively minor, but they may cause some slight differences in the effects of the institutional variables. Further research should examine provincial differences and causing reasons more closely.

Summary

This dissertation has explained theoretically why an aggregate analysis gives misleading results mostly due to uncontrolled confounding factors among the individual and institution variables. Moreover, the aggregation of data causes serious difficulties in interpreting the results. Essentially, the school effectiveness literature shows that individuals react differentially to their educational environments. In comparison to aggregate analysis, multilevel analysis avoids a number of serious methodological and conceptual difficulties. Further, using real data, the Canadian 2007 TIMSS data in British Columbia and Ontario, this study has demonstrated the bias produced by aggregating data in comparison to the improved parameter estimates produced by a multilevel analysis. The analysis show that some findings in the school effectiveness literature, specially the claimed strong and positive impact of the schools' physical resources, may not be true largely because of the inappropriate aggregation of data (Cash, 1993; Earthman & Lemasters, 1996; Tanner, 2000; Tanner & Lackney, 2006; Uline & Tschannen-Moran, 2008; Uline, Tschannen-Moran, & Wolsey, 2009). Consequently, the more advanced multilevel modeling procedure, specifically HLM, was used in this dissertation to analyze the Canadian TIMSS 2007.

In conclusion, to evaluate cross-level effects of schools, multilevel analysis of data should be used over aggregate analysis, and even over disaggregate analysis. To make confident conclusions about school effectiveness, students' characteristics should always be controlled as they have been in this study (Bosker & Witziers, 1996; Bryk & Raudenbush, 1992; Coleman, Campbell, Hobson, McPartland, Mood, Weinfeld, & York, 1966; Creemers, 1994; Rowe & Hill, 1998; Luyten, 1994; Ma, 2008; Madaus, Kellaghan, Rakow, & King, 1979; Scheerens & Bosker, 1997; Stringfield & Teddlie, 1989; Willms & Smith, 2004). Thus, a multilevel framework of schooling should be used so that the cross-level interaction effects between certain conditions in provinces, districts, schools, classrooms and teachers, and characteristics of students are all taken into consideration in analytical procedures. When multilevel analysis is not feasible, disaggregate analysis can often produce meaningful results as they did in this study. If, unfortunately, aggregate analysis is the only option, the result should not be used to infer any causal relationships involving individual students.

However, the findings in this study are based on restricted data from British Columbia and Ontario due to relatively large amounts of missing values in both school and student data. The missing values did not occur randomly and may have affected the results of the study. Therefore, researchers must design better survey questions to inquire about each student's backgrounds, especially SES. Though the results are consistent with the literature, it is better to test the same model with a full set of data on all the important variables in future studies. Finally, the adequacy of physical resources in schools assessed by school principals shows no positive impact on students' achievement. This does not necessarily mean that schools' facilities and students' achievement are not related. Rather, it means that because of improved and equalized conditions of schools, the measurement of physical resources in schools probably should be changed to the utilization of available facilities, which would more closely relate to the practice of teaching and learning. Thus, further research on the use of the physical and human resources can help administrators and policymakers ensure that school resources have been invested where they are the most effective in helping students succeed academically.

REFERENCES

21st Century School Fund. (2009). Research on the impact of school facilities on students and teachers: A summary of studies published since 2000. Retrieved from http://www.21csf.org/csf-home/Documents/ResearchImpactSchoolFacilitiesFeb2010.pdf.

Aitkin, M., & Longford, N. (1986). Statistical modeling issues in school effectiveness studies. *Journal of the Royal Statistical Society. Series A*, *149*(1), 1–43.

Alexander, K. L., & Eckland, B. K. (1975). Contextual effects in the high school attainment process. *American Sociological Review*, *40*, 402–416.

Alexander, K. L., Eckland, B. K., & Griffin, L. J. (1975). The Wisconsin model of socioeconomic achievement: A replication. *The American Journal of Sociology*, *81*(2), 324–342.

Alexander, K. L., Fennessey, J., McDill, E. L., & D'Amico, R. J. (1979). School SES influences—composition or context? *Sociology of Education*, *52*(4), 222–237.

Alwin, D. F., & Otto, L. B. (1977). High school context effects on aspirations. *Sociology of Education*, *50*, 259–273.

Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology*, *84*, 261–274.

Andrew, M., & Herb, M. (2005). Motivating boys and motivating girls: Does teacher gender really make a difference? *Australian Journal of Education*, *49*(3), 320–334.

Andrew, M., & Schwab, R. L. (1995). Has reform in teacher education influenced teacher performance? An outcome assessment of graduates of eleven teacher education programs. *Action in Teacher Education*, *17*, 43–53.

Andrews, R., & Soder, R. (1987). Principal instructional leadership and school achievement, *Educational Leadership*, *44*, 9–11.

Babad, E. (1993). Pygmalion-25 years after interpersonal expectations in the classroom. In P. D. Blanck (Ed.), *Interpersonal expectations* (pp. 125–153). Cambridge, U.K.: Cambridge University Press.

Bain, R. K., & Anderson, J. G. (1974). School context and peer influences on educational plans of adolescents. *Review of Educational Research*, *44*, 429–445.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall Inc.

Barcikowski, R. S. (1981). Statistical power with group mean as the unit of analysis. *Journal of Educational Statistics*, *6*(3), 267–285.

Barr, R., Dreeben, R., & Wiratchai, N. (1983). *How schools work*. Chicago, IL: University of Chicago Press.

Beller, M., & Gafni, N. (1996). The 1991 international assessment of educational progress in mathematics and sciences: The gender differences perspective. *Journal of Educational Psychology*, *88*(2), 365–377.

Berner, M. M. (1993). Building conditions, parental involvement, and student achievement in the District of Columbia public school system. *Urban Education*, *28*(1), 6–29.

Bidwell, C. E. (1972). Schooling and socialization for moral commitment. *Interchange, 3,* 1–26.

Bidwell, C. E., & Kasarda, J. D. (1980). Conceptualizing and measuring the effects of school and schooling. *American Journal of Education*, *88*(4), 401–430.

Blackmore, J., Bateman, D., Loughlin, J., O'Mara, J., & Aranda, G. (2011). *Research into the connection between built learning spaces and student outcomes*. East Melbourne, Victoria: Department of Education and Early Childhood Development.

Blincoe, J. M. (2008). *The age and condition of Texas high schools as related to student academic achievement*. (Unpublished doctoral dissertation). University of Texas, Austin, Texas.

Bloom, A. (2008, May 8). It's good to have a man or four about the place. *Times Educational Supplement*. Retrieved from http://www.tes.co.uk/article.aspx?storycode=2619345.

Boese, S., & Shaw, J. (2005). *New York state school facilities and student health, achievement and attendance: A data analysis report*. Healthy Schools Network.

Borg, W. R. (1979). Teacher coverage of academic content and pupil achievement. *Journal of Educational Psychology*, *71*, 635–645.

Bosker, R.J., & Witziers, B. (1996). *The magnitude of school effects, or: Does it really matter which school a student attends*? Paper presented at the annual meeting of the American Educational Research Association, New York, NY.

Bossert, S., Dwyer, D., Rowan, B., & Lee, G. (1982). The instructional management role of the principal. *Educational Administration Quarterly, 18*(3), 34–64.

Bowers, A. J., & Urick, A. (2011). Does high school facility quality affect student achievement? A two-level hierarchical linear model. *Journal of Education Finance*, *37*(1), 72–94.

Brayfield, A. H., & Crockett, W. H. (1955). Employee attitudes and employee performance. *Psychological Bulletin*, *52*, 396–424.

Bring, J. (1994). How to standardize regression coefficients. *The American Statistician*, *48*(3), 209–213.

Brochu, P., Gluszynski, T., Knighton, T. (2010). *Measuring up: Canadian results of the OECD PISA study*. Ottawa, ON: Statistics Canada.

Bronfenbrenner, U. (1976). The experimental ecology of education. *Educational Researcher*, *5*(9), 5–15.

Brookover, W. B., Beady, C., Flood, P., Schweitzer, J., & Wisenbaker, J. (1979). *Schools, social systems and student achievement: Schools can make a difference*. New York: Praeger.

Brophy, J. (1998). *Expectations in the classroom* (vol 7). New York: JAI Press.

Brophy, J., & Evertson, C. (1976). *Learning from teaching: A developmental perspective*. Boston, MA: Allyn & Bacon.

Brown, B. W., & Saks, D. H. (1981). The microeconomics of schooling. *Review of Research in Education*, *9*(1), 217–254.

Bryk, A.S., & Raudenbush, S.W. (1992). *Hierarchical Linear Models: Applications and Data Analysis Methods.* Newbury Park, CA: Sage.

Buckley, J., Schneider, M., & Shang, Y. (2004). *LAUSD School Facilities and Academic Performance*. Retrieved from http://www.edfacilities.org/pubs/LAUSD%20Report.pdf. Washington, DC: National Clearinghouse for Educational Facilities.

Bullock, C. (2007). *The relationship between school building conditions and student achievement in the middle school level of the commonwealth of Virginia*. (Unpublished doctoral dissertation). Virginal Polytechnic Institute and State University, Blacksburg, VA.

Burstein, L. (1980). The analysis of multi-level data in educational research and evaluation. *Review of Research in Education*, *8*(1), 158–233.

Burstein, L., Linn, R. L., & Capell, F. J. (1978). Analyzing multilevel data in the presence of heterogeneous within-class regressions. *Journal of Educational and Behavioral Statistics*, *3*(4), 347–383.

Busing, F. (1993). *Distribution characteristics of variance estimates in two level models*. (Unpublished manuscript). Leiden University, Department of Psychometrics and Research Methodology.

Butler, R. (1993). Effects of task- and ego-achievement goals on information seeking during task engagement. *Journal of Personality and Social Psychology, 65*(1), 18–31.

Burtless, G. E. (1996). *Does money matter? The effect of school resources on student achievement and adult success*. Washington, DC: Brookings Institution.

Calderhead, J. (1996). Teachers, beliefs, and knowledge. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 709–725). New York: Simon & Schuster.

Caprara, G. V., Barbaranelli, C., Steca, P., & Malone, P. S. (2006). Teachers' self-efficacy beliefs as determinants of job satisfaction and students' academic achievement: A study at the school level. *Journal of School Psychology*, *44*(6), 473–490.

Carnoy, M. (1971). *Family background, school inputs and students' performance in school: the case of Puerto Rico*. (Unpublished doctoral dissertation). Stanford University, School of Education, Stanford, California, USA.

Cash, C. S. (1993). *Building condition and student achievement and behavior*. (Unpublished doctoral dissertation). Virginia Polytechnic Institute and State University, Blacksburg, VA.

Cellini, S. R., Ferreira, F. V., & Rothstein, J. (2008). *The value of school facilities: Evidence from a dynamic regression discontinuity design*. Washington, DC: National Bureau of Economic Research.

Cervini, R. A. (2009). Class, school, municipal, and state effects on mathematics achievement in Argentina: A multilevel analysis. *School Effectiveness and School Improvement*, *20*(3), 319–340.

Clifton, R. A., Perry, R. P., Parsonson, K., & Hryniuk, S. (1986). Effects of ethnicity and sex on teachers' expectations of junior high school students. *Sociology of Education*, *59*(1), 58–67.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Coker, H., Medley, D. M., & Soar, R. S. (1980). How valid are expert opinions about effective teaching? *The Phi Delta Kappan, 62*(2), 131–134.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Government Printing Office.

Cooley, W., & Leinhardt, G. (1978). *The instructional dimensions study: Final report*. Pittsburgh: Learning Research and Development Center, University of Pittsburgh.

Cooper, H. (2007). *The battle over homework: Common ground for administrators, teachers, and parents*. Thousand Oaks, CA: Corwin Press.

Cooper, H. (1989). *Homework*. White Plains, NY: Longman.

Cooper, H., Lindsay, J. J., Nye, B., & Greathouse, S. (1998). Relationships among attitudes about homework, amount of homework assigned and completed, and student achievement. *Journal of Educational Psychology*, *90*, 70–83.

Cooper, H., Robinson, J. C., & Patall, E. A. (2006). Does homework improve academic achievement? A synthesis of research, 1987–2003. *Review of Educational Research*, *76*, 1–62.

Cooper, H., & Valentine, J. C. (2001). Using research to answer practical questions about homework. *Educational Psychologist*, *36*, 143–153.

Crampton, F. E. (2009). Spending on school infrastructure: Does money matter? *Journal of Educational Administration, 47*(3), 305–322.

Creemers, B. P. M. (1994). *Effective classroom*. London: Cassell.

Cronbach, L. J. (1976). *Research on classrooms and schools: Formulation of questions, design and analysis*. Stanford, CA: Stanford Evaluation Consortium.

De Jong, R., Westerhof, K. J., & Creemers, B. P. M. (2000). Homework and student math achievement in junior high schools. *Educational Research and Evaluation*, *6*, 130–157.

Denton, J. J., & Peters, W. H. (1988). *Program assessment report: Curriculum evaluation of a non-traditional program for certifying teachers*. College Station, TX: Texas A&M University.

Dunkin, M. J., & Doenau, S. J. (1980). A replication study of unique and joint contributions to variance in student achievement. *Journal of Educational Psychology*, *72*(3), 394–403.

Duncan, O. D., Haller, A. O., & Portes, A. (1968). Peer influences on aspirations: A reinterpretation. *American Journal of Sociology*, *74*, 119–137.

Duran-Narucki, V. (2008). School building condition, school attendance, and academic achievement in New York City public schools: A mediation model. *Journal of Environmental Psychology*, *28*, 278–286.

Dusek, J. B. (1985). *Teacher expectancies*. Hillsdale, NJ: Erlbaum.

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, *95*(2), 256–273.

Dweck, C. S., & Sorich, L. A. (1999). Mastery-oriented thinking. In C. R. Snyder (Ed.), *Coping: The psychology of what works*. New York: Oxford University Press.

Earthman, G. I. (2004). *Prioritization of 31 criteria for school building adequacy*. Maltimore, MD: American Civil Liberties Union Foundation of Maryland.

Earthman, G., Cash, C., & Van Berkum, D. (1995, October). *A statewide study of student achievement and behavior and school building condition*. Paper presented at the annual meeting of the Council of Education Facility Planner, International, New York.

Earthman, G. I., & Lemasters, L. K. (2009). Teacher Attitudes about classroom conditions. *Journal of Education Administration*, *47*(3), 323–335.

Earthman, G. I., & Lemasters, L. K. (1996, October). *Review of the research on the relationship between school buildings, student achievement, and student behavior*. Paper presented at the Annual Meeting of the Council of Educational Facility Planners International, Tarpon, FL.

Edmonds, R. R. (1979). Effective schools for the urban poor. *Educational Leadership*, *37*, 15–27.

Ernest, P. (1989). The knowledge, beliefs and attitudes of the mathematics teacher: A model. *Journal of Education for Teaching*, *15*, 13–34.

Ethington, C. A. (1997). A hierarchical linear modeling approach to studying college effects. In J. Smart (ed.), *Higher Education: Handbook of Theory and Research*, 12, (pp. 165–194). New York: Agathon.

Firebaugh, G. (1978). A rule for inferring individual-level relationships from aggregate data. *American Sociological Review*, *43*(4), 557–572.

Foy, P., & Olson, J. F. (2009). *TIMSS 2007 user guide for the international database*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center.

Friedman, L. (1989). Mathematics and the gender gap: A meta-analysis of recent studies on sex differences in mathematical tasks. *Review of Educational Research, 59*(2), 185–213.

Gardner, R. C., & MacIntyre, P. D. (1991). An instrumental motivation in language study: Who says it isn't effective? *Studies in Second Language Acquisition*, *13*(1), 57–72.

Goldhaber, D. (2002). The mystery of good teaching. *Education Next*, *2*(1). Retrieved from http://www.nuatc.org/articles/pdf/mystery_goodteaching.pdf.

Goldstein, H. (1995). *Multilevel statistical models* (2nd ed.). London, United Kingdom: Edward Arnold.

Goldstein, H. (1997). Methods in school effectiveness research. *School Effectiveness and School Improvement*, *8*(4), 369–395.

Goldstein, H. & Rasbash, J. (1990). *Improved approximations for multilevel models with binary responses*. London, United Kingdom: University of London.

Good, T. L., & Beckerman, T. M. (1978). Time on task: A naturalistic study in sixth-grade classrooms. *Elementary School Journal*, *78*, 193–201.

Good, T. L., Grouws, D. A., & Beckerman, T. M. (1978). Curriculum Pacing: Some Empirical Data in Mathematics. *Journal of Curriculum Studies*, *10*, 75–81.

Goodman, L. (1959). Some alternatives to ecological correlation. *American Journal of Sociology, 64*, 610–625.

Goodman, L. A. (1953). Ecological regressions and behavior of individuals. *American Sociological Review*, *18*(6), 663–664.

Gove, W. R., & Hughes, M. (1980). Reexamining the ecological fallacy: A study in which aggregate data are critical in investigating the pathological effects of living alone. *Social Forces, 58,* 1157–1177.

Greenland, S. (2001). Ecologic versus individual-level sources of bias in ecologic estimates of contextual health effects. *International Journal of Epidemiology, 30*(6), 1343–1350.

Greenland, S., & Morgenstern, H. (1989). Ecological bias, confounding, and effect modification. *International Journal of Epidemiology*, *18*, 269–274.

Greenland, S., & Robins, J. M. (1994). Ecologic studies—biases, misconceptions, and counterexamples. *American Journal of Epidemiology*, *139*, 747–760.

Hacker, A. (1997). *Money: Who has how much and why*. New York: Scribner.

Haller, A. O., & Butterworth, C. E. (1960). Peer influences on levels of occupational and educational aspirations. *Social Forces*, *38*, 289–295.

Haller, A. O., & Woelfel, J. (1972). Significant others and their expectations: Concepts and instruments to measure inter-personal influence on status aspirations. *Rural Sociology*, *37*, 591–621.

Hallinger, P. (2003). Leading Educational Change: reflections on the practice of instructional and transformational leadership. *Cambridge Journal of Education*, *33*(3), 329–351.

Hallinger, P., & Heck, R. H. (1996). Reassessing the Principal's Role in School Effectiveness: A Review of Empirical Research, 1980–1995. *Educational Administration Quarterly*, *32*(1), 5–44.

Halpern, D. F. (2000). *Sex differences in cognitive abilities* (3rd ed.). Philadelphia, PA: Lawrence Erlbaum Associates.

Handel, W. (1981). Danger of committing the ecological fallacy persists: Comment on Gove and Hughes. *Social Forces*, *60*(2), 585–588.

Hannan, M. T., & Burstein, L. (1974). Estimation from grouped observations. *American Sociological Review*, *39*(3), 374–392.

Hanushek, E. A. (2003). The failure of input-based schooling policies. *The Economic Journal*, *113*(485), F64–F98.

Hanushek, E. A. (1997). Assessing the effects of school resources on student performance: An update. *Educational Evaluation and Policy Analysis*, *19*(2), 141–164.

Hanushek, E. (1996). Measuring investment in education. *Journal of Economic Perspectives*, *10*(4), 9–30.

Hanushek, E. A. (1986). The economics of schooling: production and efficiency in public schools. *Journal of Economic Literature*, *49*(3), 1141–1177.

Hanushek, E. A., Rivkin, S. G., & Taylor, L. L. (1996). Aggregation and the estimated effects of school resources. *Review of Economics and Statistics*, *78*(4), 611–627.

Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (1998). *Teachers, schools, and academic achievement* (working paper No. 6691). Cambridge, MA: National Bureau of Economic Research.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.

Hattie, J., & Clinton, J. (2001). The assessment of teachers. *Teaching Education*, *12*, 279–300.

Hauser, R. M. (1970). Context and consex: A cautionary tale. *American Journal of Sociology*, *75*(4), 645–664.

Hauser, R. M. (1971). *Socioeconomic background and educational performance*. Washington, DC: American Sociological Association.

Hauser, R. M., Sewell, W. H., & Alwin, D. F. (1976). High school effects on achievement. In W. H. Sewell & R. M. Hauser (Eds.), *Schooling and achievement in American society* (pp. 309–341). New York: Academic Press.

Hedges, L.V., Laine, R., & Greenwald, R. (1994). Does money matter? A meta-analysis of studies of the effects of differential school inputs on student outcomes. *Educational Researcher, 23*(3), 5–14.

Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science, 269*(5220), 41–45.

Henry, G. T., Fortner, C. K., & Thompson, C. L. (2010). Targeted funding for educationally disadvantaged students: A regression discontinuity estimate of the impact on high school student achievement. *Educational Evaluation and Policy Analysis*, *32*(2), 183–204.

Heppell, S., Chapman, C., Millwood, R., Constable, M., & Furness, J. (2004). *Building learning futures: A research project at ultra lab within CABE/RIBA "Building Futures" programme*. Chelmsford, UK: Ultralab.

Herriott, R. E. (1963). Some determinants of educational aspiration. *Harvard Educational Review*, *33*, 157–177.

Heschong Mahone Group. (1999). *Daylighting in schools: An investigation into the relationship between daylighting and human performance*. Fair Oaks, CA: Heschong Mahone Group.

Holmlund, H., & Sund, K. (2005). Is the gender gap in school performance affected by the sex of the teacher? *Labour Economics, 15*(1), 37–53**.**

Hox, J. J. (2002). *Multilevel analysis: Techniques and applications*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Hughes, S. M. (2006). *The relationship between school design variables and student achievement in a large urban Texas school district*. (Unpublished doctoral dissertation). Baylor University, Waco, TX.

Human Resources and Skills Development Canada, Council of Ministers of Education, Canada and Statistics Canada. 2010. *The performance of Canada's youth in reading, mathematics and science. PISA 2009 first results for Canadians Aged 15.* Ottawa, Canada: Minister of Industry.

Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin*, *104*(1), 53–69.

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin*, *107*(2), 139–155.

Jackson, P. (1968). *Life in classrooms*. New York: Holt, Rinehart & Winston.

Jencks, C.S., Smith, M., Ackland, H., Bane, M.J., Cohen, D., Ginter, H., Heyns, B., & Michelson, S. (1972). *Inequality: A reassessment of the effect of the family and schooling in America*. New York: Basic Books.

Jussim, L., & Harber, K. D. (2005). Teacher expectations and self-fulfilling prophecies: Knowns and unknowns, resolved and unresolved controversies. *Personality and Social Psychology Review*, *9*(2), 131–155.

Jussim, L., Smith, A., Madon, S., & Palumbo, P. (1998). Teacher expectations. In J. Brophy (Ed.), *Advances in research on teaching* (vol 7, pp. 1–48). Greenwich, CT: JAI.

Kandel, D., & Lesser, G. S. (1969). Educational plans of adolescents. *American Sociological Review*, *34*, 213–223.

Kaplan, A., & Maehr, M. L. (1999). Achievement goals and student well-being. *Contemporary Educational Psychology*, *24*(4), 330–358.

Keith, T. Z. (1986). *Homework*. West Lafayette: Kappa Delta Pi.

Kelly, H. H. (1952). Two functions of reference groups. In G. E. Swanson, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in Social Psychology* (pp. 410–414). New York: Holt.

Kim, K. S. (1990). *Multilevel data analysis: A comparion of analytical alternatives*. Unpublished doctorate thesis. Los Angeles, CA: University of California.

King, G. (1997). *A solution to the ecological inference problem: Reconstructing individual behavior from aggregate data*. Princeton, NJ: Princeton University Press.

Kish, L. (1965). *Survey sampling*. New York: Wiley.

Klees, S. J. (1974). *Television and other determinants of scholastic achievement in Mexican secondary education*. (Unpublished doctoral dissertation). Stanford University, Institute for Communication Research, Stanford, CA.

Klitgaard, R. E., & Hall, G. R. (1974). Do schools make a difference? *Journal of Human Resources*, *74*, 90–106.

Kousser, J. M. (2001). Ecological inference from Goodman to King. *Historical Methods*, *34*(3), 101–126.

Kreft, Ita G.G. (1996). *Are multilevel techniques necessary? An overview, including simulation studies*. Los Angeles: California State University.

Krueger, A. B. (1999). Experimental estimates of education production functions. *The Quarterly Journal of Economics*, *114*(2), 497–532.

Kulik, J. A. (1992). *An analysis of the research on ability grouping: Historical and contemporary perspectives*. Storrs, CT: National Research Center on the Gifted and Talented, University of Connecticut.

Kulik, C. C., & Kulik, J. A. **(**1982). Effects of ability grouping on secondary school students: A meta-analysis of evaluation findings. *American Educational Research Journal*, *19*(3), 415–428.

Laffaldano, M. T., & Muchinsky, P. M. (1985). Job satisfaction and job performance: A meta-analysis. *Psychological Bulletin*, *97*, 251–273.

Lam, Y.H. R., Tse, S. K., Lam, J. W.I., & Loh, E. K.Y. (2010). Does the gender of the teacher matter in the teaching of reading literacy? Teacher gender and pupil attainment in reading literacy in Hong Kong. *Teaching and Teacher Education*, *26*(4), 754–759.

Laurie, R. (2007). *Setting them up to fail? Excellent school marks don't necessarily lead to excellent exam marks*. Retrieved from http://www.aims.ca/site/media/aims/Gradeinflationvfinal_cc.pdf. Halifax NS: Atlantic Institute for Market Studies.

Lee, V. E., & Bryk, A. S. (1989). A multilevel model of the social distribution of high school achievement. *Sociology of Education*, *62*(3), 172–192.

Lewis, M. (2000). *Where children learn: Facility condition and student test performance in Milwaukee public schools*. Scottsdale, AZ: Council of Education Facility Planners.

Little, R., & Rubin, D. B. (1989). The analysis of social science data with missing values. *Sociological Methods Research*, *18*(2–3), 292–326.

Lomas, C. & Oblinger, D.J. (2005), *Student practices and their impact on learning spaces*. In D.G. Oblinger (Ed.), *Learning Spaces*, (pp. 5.1–5.11). Washington, DC: Educause.

Lortie, D. (1975). *Schoolteacher: A sociological study*. Chicago: University of Chicago Press.

Luyten, H. (1994). *School effects: Stability and malleability*. Enschede: University of Twente.

Ma, X., & Klinger, D. (2000). Hierarchical linear modeling of student and school effects on academic achievement. *Canadian Journal of Education*, *25*(1), 41–55.

Ma, X. (2008). Using multilevel modeling to investigate school effects. In A. A. O'Connell & D. B. McCoach (Eds.), *Multilevel modeling of educational data* (pp. 59–110). Charlotte, NC: Information Age Publishing.

Madaus, G. F., Kellaghan, T., Rakow, E.A., & King, D. J. (1979). The sensitivity of measures of school effectiveness. *Harvard Educational Review*, *49*, 207–230.

Marjoribanks, K. (2005). Family background, adolescents' educational aspirations, and Australian young adults' educational attainment. *International Education Journal*, *6*(1), 104–112.

Marks, J., Cox, C., & Pomian-Srzednicki, M. (1985). *Standards in English schools second report: An analysis of the examination results of secondary schools in England for 1982 and comparisons with 1981*. Great Britain: National Council for Educational Standards.

Marks, J., Cox, C., & Pomian-Srzednicki, M. (1983). *Standards in English schools: An analysis of the examination results of secondary schools in England for 1981*. Great Britain: National Council for Educational Standards.

McDill, E. L., Meyers, Jr., E. D., & Rigsby, L. C. (1967). Institutional effects on the academic behavior of high school students. *Sociology of Education*, *40*, 181–189.

McDill, E. L., & Rigsby, L. C. (1973). *Structure and process in secondary schools: The impact of educational climates*. Baltimore: The Johns Hopkins University Press.

McDill, E. L., Rigsby, L. C., & Meyers, Jr., E. D. (1969). Educational climates of high schools: Their effects and sources. *American Journal of Sociology*, *74*, 567–586.

McGuffey, C. W., & Brown, C. L. (1978). The impact of school building age on school achievement in Georgia. *Council of Educational Facility Planners Journal*, *16*, 6–9.

Meyer, R. H. (2001). *Estimation of teacher and school performance in the Denver public schools: A feasibility study*. Madison, WI: Wisconsin Center for Education Research, University of Wisconsin-Madison.

Meyer, J. H. (1970). High school effects on college intentions. *American Journal of Sociology*, *76*, 59–70.

Miller, R. B., Greene, B. A., Montalvo, G. P., Ravindran, B., & Nichols, J. D. (1996). Engagement in academic work: The role of learning goals, future consequences, pleasing others, and perceived ability. *Contemporary Educational Psychology*, *21*(4), 388–422.

Mischel, W. (1977). The interaction of person and situation. In D. Magnusson & N. S. Endler (Eds.), *Personality at the crossroads. Current issues in interactional psychology* (pp. 333–352). Hillsdale, NJ: Erlbaum.

Morgenstern, H., & Thomas, D. (1993). Principles of study design in environmental epidemiology. *Environmental Health Perspectives*, *101*(4), 23–38.

Mueller, C. W., & Parcel, T. L. (1981). Measures of socioeconomic status: Alternatives and recommendations. *Child Development*, *52*, 13–30.

Muhlenbruck, L., Cooper, H., Nye, B., & Lindsay, J. J. (2000). Homework and achievement: Explaining the different strengths of relation at the elementary and secondary school levels. *Social Psychology of Education*, *3*, 295–317.

Murnane, R. J., & Phillips, B. R. (1981). Learning by doing, vintage, and selection: Three pieces of the puzzle relating teaching experience and teaching performance. *Economics of Education Review*, *1*(4), 453–465.

Mwamwenda, T. S., & Mwamwenda, B. B. (1989). Teacher characteristics and pupils' academic achievement in Botswana primary education. *International Journal of Educational Development*, *9*(1), 31–34.

Nair, P., & Fielding, R. (2005). *The language of school design: Design patterns for 21st century schools*. Minneapolis, MN: Designshare.

Nash, R. (2003). Is the school composition effect real?: A discussion with evidence from the uk PISA data. *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice*, *14*(4), 441–457.

Nelson, J. I. (1972). High school context and college plans: The impact of social structure on aspiration. *American Sociological Review*, *37*, 143–148.

Nespor, J. (1987). The role of beliefs in the practice of teaching. *Journal of Curriculum Studies*, *19*, 317–328.

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, *26*(3), 237–257.

O'Neill, D. J. (1999). *The impact of school facilities on student achievement, behavior, attendance, and teacher turnover rate at selected Texas middle schools in region XIII ESC*. (Unpublished doctoral dissertation). Texas A&M University, College Station, TX.

Oakes, J. M. (2009). Commentary: Individual, ecological and multilevel fallacies. *International Journal of Epidemiology*, *38*, 361–368.

Organization for Economic Cooperation & Development. (2000). *The appraisal of investments in educational facilities*. Paris: OECD.

Palardy, G. J., & Rumberger, R. W. (2008). Teacher effectiveness in first grade: The importance of background qualifications, attitudes, and instructional practices for student learning. *Educational Evaluation and Policy Analysis, 30*(2), 111–140.

Paterson, L. (1991). An introduction to multilevel modelling. In S. W. Raudenbush & J. D. Willms (Eds.), *Schools, classrooms, and pupils* (pp. 13–25). San Diego, CA: Academic Press, Inc.

Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: a review of reporting practices and suggestions for improvement. *Review of Educational Research*, 74(*4*), 525–556.

Piantadosi, S. (1994). Invited commentary: Ecologic biases. *American Journal of Epidemiology*, *139*, 761–764.

Picus, L. O. (2001). *In search of more productive schools: A guide to resource allocation in education*. Eugene, OR: ERIC Clearinghouse on Educational Management, University of Oregon.

Picus, L. O., Marion, S. F., Calvo, N., & Glenn, W. J. (2005). Understanding the relationship between student achievement and the quality of educational facilities: Evidence from Wyoming. *Peabody Journal of Education*, *80*(3), 71–95.

Pintrich, P. R. (2000). An achievement goal theory perspective on issues in motivation terminology, theory, and research. *Contemporary Educational Psychology*, *25*(1), 92–104.

Piper, J. E. (2004). *Handbook of facility assessment*. New York: Marcel Dekker Inc.

Porter, A . C., & Freeman, D . J. (1986). Professional orientations: An essential domain for teacher testing. *Journal of Negro Education*, *55*, 284–292.

Raudenbush, S. W. (1984). Magnitude of teacher expectancy effects on pupil IQ as a function of the credibility of expectancy induction: A synthesis of findings from 18 experiments. *Journal of Educational Psychology*, *76*(1), 85–97.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Newbury Park, CA: Sage.

Raudenbush, S. W., & Bryk, A. S. (1986). A hierarchical model for studying school effects. *Sociology of Education*, *59*, 1–17.

Raudenbush, S. W., & Wilms, J. D. (1995). The estimation of school effects. *Journal of Educational Behavioral Statistics*, *20*, 307–335.

Reynolds, D., Teddlie, C., Creemers, B., Scheerens, J., & Townsend, T. (2000). An introduction to school effectiveness research. In C. Teddlie & D. Reynolds (Eds.), *The international handbook of school effectiveness research*. London: Falmer Press.

Reynolds, A. J., & Walberg, H. J. (1992). A process model of mathematics achievement and attitude. *Journal for Research in Mathematics Education, 23*(4), 306–328.

Richards, J. (2008). Closing the Aboriginal/non-Aboriginal education gaps. *C.D. Howe Institute Backgrounder*, 116.

Richardson, V. (1994). The consideration of teachers' beliefs. In V. Richardson (Ed.), *Teacher change and the staff development process* (pp. 90–108). New York: Teachers College Press.

Rivers, J.C., & Sanders, W.L. (2000, May). *Teacher quality and equity in educational opportunity: findings and policy implications*. Paper presented at the Hoover/PRI Teacher Quality Conference, Palo Alto, CA.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, *73*(2), 417–458.

Roberts, L. W. (2009). Measuring school facility conditions: An illustration of the importance of purpose. *Journal of Educational Administration, 47*(3), 368–380.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, *15*, 351–357.

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: evidence from panel data. *American Economic Review*, *94*(2), 247–252.

Rodriguez, G. & Goldman, N. (1995). An assessment of estimation procedures for multilevel models with binary responses. *Journal of the Royal Statistical Society. Series A (Statistics in Society), 158*(1), 73–89.

Roeser, R. W., Midgley, C., & Urdan, T.C. (1996). Perceptions of the school psychological environment and early adolescents' psychological and behavioral functioning in school: The mediating role of goals and belonging. *Journal of Educational Psychology*, *88*(3), 408–422.

Rogosa, D. (1978). Politics, process, and pyramids. *Journal of Educational and Behavioral Statistics*, *3*, 79–86.

Rosenthal, R. (1991). Teacher expectancy effects: A brief update 25 years after the Pygmalion experiment. *Journal of Research in Education*, *1*(1), 3–12.

Rosenholtz, S. J., & Bassler, O. (1986). Organizational conditions of teacher learning. *Teaching and Teacher Education*, *2*(2), 91–104.

Rosenholtz, S. J., & Smylie, M. A. (1984). Teacher Compensation and Career Ladders. *The Elementary School Journal*, *85*(2), 149–166.

Rowe, K. J., & Hill, P. W. (1998). Modeling educational effectiveness in classrooms: The use of multi-level structural equations to model students' progress. *Educational Research and Evaluation*, *4*(4), 307–347.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(*3*), 581–592.

Rutter, M., Maughan, B., Mortimore, P., & Ouston, J. (1979). *Fifteen thousand hours: Secondary schools and effects on children*. Boston: Harvard University Press.

Saha, L. J. (1983). Social structure and teacher effects on academic achievement: A comparative analysis. *Comparative Education Review*, *27*(1), 69–88.

Sanders, W. L., & Horn, S. P. (1998). Research findings from the Tennessee Value-Added Assessment System (TVAAS) Database: Implications for educational evaluation and research. *Journal of Personnel Evaluation in Education*, *12*(3), 247–256.

Sanders, W. L., & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement. Research Progress Report*. Knoxville: University of Tennessee Value-Added Research and Assessment Center.

Scheerens, J., & Bosker, R. J. (1997). *The foundations of educational effectiveness*. Oxford/New York/Tokyo: Pergamon.

Schneider, M. (2003). *Linking school facility conditions to teacher satisfaction and success.* Washington, DC: National Clearinghouse for Educational Facilities.

Selvin, H. C. (1958). Durkheim's suicide and problems of empirical research. *American Journal of Sociology*, *63*, 607–619.

Sewell, W. H., Haller, A. O., & Ohlendorf, G. W. (1970). The educational and early occupational attainment process: Replication and revision. *American Sociological Review*, *35*, 1014–1027.

Sewell, W. H., Haller, A. O., & Portes, A. (1969). The educational and early occupational attainment process. *American Sociological Review, 34*, 82–92.

Sewell, W. H., & Hauser, R. M. (1975). *Education, occupation, and earnings. achievement in the early career*. New York: Academic Press Inc.

Sheets, M. E. (2009). *The relationship between the condition of school facilities and certain educational outcomes, particularly in rural public high schools in Texas*. (Unpublished doctoral dissertation). Texas Technology University, Lubbock, TX.

Shively, W. P. (1969). Ecological inference: The use of aggregate data to study individuals. *The American Political Science Review*, *63*(4), 1183–1196.

Simpson, R. L. (1962). Parental influence, anticipatory socialization, and social mobility. *American Sociological Review*, *27*, 517–522.

Sirin, S. R. (2005). SocioEconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, *75*(3), 417–453.

Sowell, T. (2009). *The housing bottom and bust*. New York, NY: Basic Books.

Stevenson, K. R. (2001). *The relationship of school facilities conditions to selected student academic outcomes: A study of South Carolina public school*s. Columbia, SC: SC Education Oversight Committee.

Stringfield, S., & Teddlie, C. (1989). The first three phases of the Louisiana School Effectiveness Study. In B. Creemers, T. Peters, & D. Reynolds (Eds.), *School effectiveness and school improvement: Proceedings of the second international congress* (pp.281–294). Amsterdam, Netherlands: Swets and Zeitlinger.

Stumpf, H. (1995). Gender differences in performance on tests of cognitive abilities: Experimental design issues and empirical results. *Learning and Individual Differences*, *7*(4), 275–287.

Subramanian, S. V., Jones, K., Kaddour, A., & Krieger, N. (2009). Revisiting Robinson: The perils of individualistic and ecologic fallacy. *International Journal of Epidemiology*, *38*(2), 342–360.

Tanner, C. K. (2009). Effects of school design on student outcomes. *Journal of Educational Administration*, *47*(3), 376–394.

Tanner, C. K. (2000). The influence of school architecture on academic achievement. *Journal of Educational Administration*, *38*(4), 309–330.

Tanner, C. K., & Lackney, J.A. (2006). The physical environment and student achievement in elementary schools. In C. K. Tanner & J. A. Lackney (Eds.), *Educational facilities planning: Leadership, architecture, and management* (pp. 266–294). Boston, MA: Pearson Education.

Tate, W. F. (1997). Race-ethnicity, SES, gender, and language proficiency trends in mathematics achievement: An update. *Journal for Research in Mathematics Education*, *28*(6), 652–679.

Tate, R. & Wongbundhit, Y. (1983). Random versus nonrandom coefficient models for multilevel analysis. *Journal of Educational Statistics*, *8*, 103–120.

Thomas, E., & Gross, D. (2002, July 23). Taxing the super rich. *Newsweek*, p.38.

Townsend, T. (2007). *International handbook of school effectiveness and improvement: Review, reflection and reframing*. Springer.

Trautwein, U. (2007). The homework–achievement relation reconsidered: Differentiating homework time, homework frequency, and homework effort. *Learning and Instruction*, *17*(3), 372–388.

Trautwein, U., & Köller, O. (2003). The relationship between homework and achievement: still much of a mystery. *Educational Psychology Review*, *15*, 115–145.

Trautwein, U., Köller, O., Schmitz, B., & Baumert, J. (2002). Do homework assignments enhance achievement? A multilevel analysis in 7th grade mathematics. *Contemporary Educational Psychology*, *27*, 26–50.

Uline., C., & Tschannen-Moran., M. (2008). The walls speak: The interplay of quality facilities, school climate, and student achievement. *Journal of Educational Administration*, *46*(1), 55–73.

Uline, C., Tschannen-Moran, M., & Wolsey, T. D. (2009). The walls still speak: The stories occupants tell. *Journal of Educational Administration*, *47*(3), 400–426.

Utman, C. H. (1997). Performance effects of motivational state: A meta-analysis. *Personality and Social Psychology Review*, *1*(2), 170–182.

Vance, V. S., & Schlechty, P. C. (1982). The distribution of academic ability in the teaching force: Policy implications. *The Phi Delta Kappan*, *64*(1), 22–27.

Van Der Leeden, R. & Busing, F. (1994). *First iteration versus IGLS/RIGLS estimates in two-level modesl: A Monte Carlo study with ML3*. Leiden: Department of Psychometrica and Research Methodology, Leiden University.

Verstegen, D. A., & King, R. A. (1998). The relationship between school spending and student achievement: A review and analysis of 35 years of production function research. *Journal of Education Finance*, *24*(2), 243–262.

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin*, *117*(2), 250–270.

Wakefield, J., & Salway, R. (2001). A statistical framework for ecological and aggregate studies. *Journal of Royal Statistics Society Services A*, *164*, 119–137.

Walberg, H. J. (1991). Does homework help? *The School Community Journal*, *1*, 13–15.

Warwich, D. P., & Jatoi, H. (1994). Teacher gender and student achievement in Pakistan. *Comparative Education Review, 38*(3), 377–399.

Wayne, A. J., &Youngs, P. (2003). Teacher characteristics and student achievement gains: a review. *Review of Educational Research*, *73*(1), 89–122.

Webb, N. M. (1977). *Learning in individual and small group settings*. Stanford, CA: Stanford University School of Education.

Webster, W. J., Mendro, R. L., Orsak, T. H., & Weerasinghe, D. (1996, April). *The applicability of selected regression and hierarchical linear models to the estimation of school and teacher effects*. Paper presented at the annual meeting of the National Council on Measurement in Education, New York.

Wei, Y., Clifton, R. A., Roberts, L. W. (2011). School resources and the academic achievement of Canadian students. *Alberta Journal of Educational Research*, *57*(4), 460–478.

White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin*, *91*(3), 461–481.

Wiles, C. A. (1992). Investigating gender bias in the evaluations of middle school teachers of mathematics. *School Science and Mathematics*, *92*(6), 295–298.

Willms, J. D. & Smith, T. (2004). *A manual for conducting analyses with data from TIMSS and PISA*. Retrieved from http://www.datafirst.uct.ac.za/wiki/images/e/e7/TIMSS_1995-99_Manual.pdf.

Wright, S. P. Horn, S. P., & Sanders, W. L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, *11*, 57–67.

Woelfel, J., & Haller, A. O. (1971). Significant others, the self-reflexive act and the attitude formation process. *American Sociological Review*, *36*, 74–87.

Wyne, M. D., & Stuck, G. B. (1982). Time and learning: Implications for the classroom teacher. *Elementary School Journal*, *83*(1), 67–75.

Xu, J. (2005). Purposes for doing homework reported by middle and high school students. *The Journal of Educational Research*, *99*, 46–55.