This is to certify that the

dissertation entitled

Access to Eighth-Grade Algebra:
A Bayesian, Multilevel Analysis

presented by

Yuk Fai Cheong

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Measurement and
Quantitative Methods

Major professor

Date 2-3-97

ACCESS TO EIGHTH-GRADE ALGEBRA: A BAYESIAN, MULTILEVEL ANALYSIS

By

Yuk Fai Cheong

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Counseling, Educational
Psychology and Special Education

1997

ABSTRACT

ACCESS TO EIGHTH-GRADE ALGEBRA: A BAYESIAN, MULTILEVEL ANALYSIS

By

Yuk Fai Cheong

This dissertation addressed two research objectives: one substantive, the other methodological. The first objective was to examine what school- and state-level factors may influence a public school's decision to offer eighth-grade algebra for high school credits. The analysis employed the data collected under the 1992 Trial State Assessment Program (TSAP) in mathematics of the National Assessment of Educational Progress (NAEP), and the state profile statistics compiled by the Council of Chief State School Officers (CCSSO) (CCSSO, 1993). The second, related objective was to develop and evaluate a fully Bayesian, multilevel approach that enables the study of public schools as distributors of learning opportunities in advanced mathematics.

The Bayesian, multilevel approach developed by Zeger and Karim (1991) was implemented via the Gibbs sampler (Gemen & Gemen, 1984; Gelfand & Smith, 1990). The algorithm was coded using the Interactive Matrix Language of the Statistical Analysis Systems (SAS/IML) (SAS Institute, Inc., 1989). The result of a small simulation study, which generated and analyzed data sets having similar structure to that of the TSAP, showed that the SAS/IML code performed well and the algorithm yielded reasonable inferences.

The tested code was used to fit two models studying the distribution of learning opportunities in mathematics. The results of the first unconditional model reveal

significant state-to-state variation in the likelihood of the offering of eighth-grade algebra by public schools. The findings of the second model suggest that schools serving minority students and students of low social-economic status (SES), small schools, schools located in rural settings and states that spend less on education are comparatively less likely than other schools to offer algebra. The implication is that size, composition, and location of a school are linked to inequality in access to this educational resource. In these ways, the schooling system reinforces social, ethnic and geographic inequalities regarding the opportunities available for eighth-grade students to study algebra for high school credits.

This dissertation is dedicated to the

loving memory of my grandmom, Ah Mah.

ACKNOWLEDGMENTS

This research was supported by a grant from the American Educational Research Association which receives funding for its "AERA Grants Program" from the National Science Foundation and the National Center for Education Statistics (U.S. Department of Education) under NSF Grant #RED-9255437. I wish to thank Dr. Richard Shavelson and Ms. Jeanie Oakes for their help in the grant application process and their continued encouragement. Opinions expressed in this work reflect only those of the author and do not necessarily reflect those of the granting agencies.

I wish to express my gratitude to my dissertation advisor, Dr. Stephen Raudenbush, for his guidance and support from the initial stage to the completion of this dissertation, for his mentoring and the many invaluable teaching and research opportunities he has given me over the years. I wish to also express my appreciation to my doctoral committee members, Dr. Betsy Becker, Dr. Aaron Pallas, and Dr. James Stapleton, whose availability, support, and critical analysis were exceptional. I am thankful to Dr. Michael Seltzer at the University of California, Los Angeles, for generously sharing his expertise in Gibbs sampling.

I also appreciate the support of my colleagues involved with the Longitudinal and Multilevel Methods Project. In particular, I am grateful to our project manager, Marcy

Wallace, for her encouragement. Special thanks to Dr. Randall Fotiu, Dr. Rafa Kasim, and Mengli Yang for their assistance in programming, and to Todd D. Harcek for reading and commenting on the manuscript.

Finally, I am especially appreciative of the love, understanding, and patience given to me by my parents, sisters, brother, and in-laws throughout the pursuit of my graduate education.

TABLE OF CONTENTS

LIST OF TABLES x

LIST OF FIGURES xi

CHAPTER 1
INTRODUCTION 1
1.1 Objectives of the Study 1
1.2 Substantive Goals 1
1.3 Methodological Goals 3
1.4 Overview of the Study 5

CHAPTER 2
BACKGROUND AND SIGNIFICANCE 7
2.1 Introduction 7
2.2 The Substantive Inquiry 7
2.2.1 Background 7
2.2.2 Relevant Literature 11
2.2.3 Sources of School-to-School Variation in the Availability of Eighth-Grade Algebra 13
2.2.4 Sources of State-to-State Variation in the Availability of Eighth-Grade Algebra 16
2.2.5 Significance of the Substantive Inquiry 17
2.3 The Methodological Work 19
2.3.1 Motivations 19
2.3.2 An Illustrative Example 19
2.3.3 The Maximum Likelihood Approach 24
2.3.4 The Approximate Maximum Likelihood Approach\Penalized Quasi-likelihood Approach 29
2.3.5 The Fully Bayesian Approach 32
2.3.6 Appropriateness of the Fully Bayesian approach 37
2.3.7 Usefulness of the Bayesian Approach to Policy Research 38

CHAPTER 3
BAYESIAN ESTIMATION OF GENERALIZED LINEAR MODELS WITH RANDOM EFFECTS VIA THE GIBBS SAMPLER 39
3.1 Introduction 39
3.2 Logic of the Gibbs Sampler 39
3.3 Implementation of the Gibbs Sampler 42
3.3.1 Full Conditional Distribution of γ--$p(\gamma|T,\boldsymbol{u},y)$ 43
3.3.2 Full Conditional Distribution of T--$p(T|\gamma,\boldsymbol{u},y)$ 47
3.3.3 Full Conditional Distribution of $\boldsymbol{u}$--$p(\boldsymbol{u}|\gamma,T,y)$ 48
3.3.4 Starting Values and Convergence Diagnostics 50
3.4 A Simulation Study and Results 57

CHAPTER 4
BAYESIAN HIERARCHICAL ANALYSES OF ACCESS TO EIGHTH-GRADE ALGEBRA 62
4.1 Introduction 62
4.2 Data, Procedures, and Measures 62
4.2.1 Data Description 62
4.2.2 Procedures 63
4.2.3 Measures 65
4.3 Results 73
4.3.1 Unconditional Model 73
4.3.2 Between-school/Within-state and Between-State Model 80
4.4 Discussion and Implications 87

CHAPTER 5
SUMMARY AND CONCLUSIONS 92
5.1 Summary and Conclusions 92
5.2 Future Research Needs 93
5.2.1 Substantive Inquiry 93
5.2.2 Methodological Development 95

APPENDIX A

SAS\IML Code for the Gibbs Sampler 97

APPENDIX B

Approval Letter from the U.S. Department of Education
for Accessing Individually Identifiable Survey Data Base 117

APPENDIX C

Approval Letter from the University Committee
on Research Involving Human Subjects 118

LIST OF REFERENCES 119

LIST OF TABLES

Table 2.1: Defining Features for the Three Major Approaches 36

Table 3.1: CODA Output for Geweke Convergence Diagnostic 52

Table 3.2: CODA Output for Raftery and Lewis Convergence Diagnostic 56

Table 3.3: Results of Simulation Study 60

Table 4.1: Descriptive Statistics 72

Table 4.2: Results of the Unconditional (No Predictors) Model 74

Table 4.3: Results of the Within- and Between-State Model 81

LIST OF FIGURES

Figure 3.1 Illustration of Rejection Sampling (Taken from Gelman et al., 1995, p. 304)45

Figure 3.2 Geweke's Convergence Diagnostic 54

Figure 4.1. Logits Plotted Against the Midpoints of the Categories of Percentages of Minority Enrollment 67

Figure 4.2: Plot of Logits Against Midpoints of Grouped Categories of Enrollment ... 70

Figure 4.3: Marginal Posterior Distribution of τ (The Unconditional Model) 76

Figure 4.4a Distribution of Predicted Probabilities of Offering Algebra for the First Half of Individual States 78

Figure 4.4b Distribution of Predicted Probabilities of Offering Algebra for the Second Half of Individual States 79

Figure 4.5: Marginal Posterior Distribution of τ (The Conditional Model) 86

CHAPTER 1

INTRODUCTION

1.1 Objectives of the Study

This dissertation has two related objectives, one substantive, and the other methodological. The first objective is to describe and study unequal learning opportunities in advanced mathematics in Grade 8 among schools across 42 states[1]. This objective inspires the development and the evaluation of a fully Bayesian treatment of random effects models with binary outcomes, which constitutes the second, methodological objective of this dissertation.

1.2 Substantive Goals

The substantive purpose of the work is to investigate school and state influences on public schools' decisions to offer eighth-grade algebra for high school credits. Eighth-grade algebra is often a "gatekeeping" course that allows students access to advanced high school mathematics. Data collected under the 1992 TSAP of NAEP, and information made available by CCSSO (CCSSO, 1993) are combined to relate school- and state-level factors to the availability of an algebra course. Inquiry into the curriculum distinctions in mathematics among schools across the forty-two states can help us understand the interschool and interstate variation, and some of its sources, in the stratification of learning opportunities in the middle grades mathematics curriculum. It can help us

---

[1]There were a total of 41 states and 1 territory in the sample. For brevity, they will be referred to as states in this dissertation.

evaluate the effect of eighth-grade algebra on student achievement by offering a more thorough understanding of the selection process. The analysis can inform discussion on school- and state-level policies related to equity and funding. Furthermore, the likelihood of course availability can perform as an opportunity-to-learn indicator (Oakes, 1989; Pelgrum, Voogt, & Plomp, 1995; Porter, 1991) in monitoring the performance of educational systems.

The study draws on perspectives primarily from Sørenson and Hallinan's (1977) conception and Gamoran's (1987) model of opportunities for learning. Raudenbush, Fotiu, Cheong and Ziazi's (1996) study on inequality in access to educational resources, and Mullis, Jenkin, and Johnson's (1994) NAEP 1992 report on school effectiveness pertaining to mathematics education, laid the groundwork for this analysis. Major empirical work that guides the formulation of various research hypotheses includes studies on course availability done by Becker (1990), MacIver and Epstein (1995), Monk and Haller (1993), Oakes (1990), and Useem (1992).

At the school level, relationships between school composition, setting, size and the likelihood of the offering of algebra for high school credits are examined. Specifically, the analysis evaluates the hypotheses that schools 1) located in rural settings, 2) with greater percentages of African and Hispanic Americans, 3) of lower SES, and 5) with smaller grade 8 enrollment are less likely to offer high school algebra. At the state level, the study examines 1) whether the availability of the course varies with state poverty levels, as indicated by the percent of children in poverty, 2) whether state educational expenditures are related to how likely the algebra course is to be offered, and

3) whether state educational expenditures moderate the relationships between a school's racial and SES composition and the likelihood that it offers algebra. Put differently, the last question asks if there is an interaction effect between state educational spending and school racial and SES composition on opportunities to learn. It is hypothesized that schools located in states with higher poverty rates, on average, are less likely to offer the advanced course. In addition, greater state educational investment efforts are associated with a greater likelihood of offering algebra. The last state-level hypothesis postulates that the association between school racial and social composition effects and the probability of offering algebra depends on how much money states allocate per student.

1.3 Methodological Goals

The inquiry has motivated the development and evaluation of a fully Bayesian analytical approach which accommodates 1) the binary nature of the outcome variable of the offering of the algebra course, 2) the hierarchical or nested design of the TSAP in mathematics, schools nested within states and territories, and 3) the need to incorporate the uncertainty that arises about the variances at the state-level. By representing the state-level variance as the investigator's uncertainty about the processes that produce it, the hierarchical framework enables us to study state-level heterogeneity, the relationships between state-level predictors and outcomes (Raudenbush, Cheong, & Fotiu, 1995), and not to treat the states as fixed strata.

The fully Bayesian estimation approach has several advantages over two other major approaches, the maximum likelihood (ML) (e.g., Karim, 1991) and the

approximate maximum likelihood (AML) (e.g., Goldstein, 1991) or the closely related penalized quasi-likelihood (PQL) approaches (e.g., Breslow & Clayton, 1993). First, relative to AML\PQL, Bayes estimates of the level-2 variances are less biased (Breslow & Clayton, 1993; Rodriguez & Goldman, 1995). Second, relative to ML and AML\PQL, Bayes inference about any parameter allows full assessment of uncertainty in other parameters in the same model. Thus, unlike the other two approaches, Bayes estimates of the regression parameters incorporate uncertainty of the estimate of variance. Lastly, the Bayesian estimation supplies researchers with fuller inferential information by providing the entire posterior distributions of estimates, rather than point and interval estimates alone. This approach is useful: given the modest number of states, the posterior distribution of the level-2 variance is likely to be skewed and inferences based solely on the variance of the estimate can be misleading.

Zeger and Karim (1991) developed an algorithm for this approach and implemented it via the Gibbs sampler (Gemen & Gemen, 1984; Gelfand & Smith, 1990). They showed with a simulation that the algorithm yielded reasonable inferences in finite sample cases in which the number of clusters was large relative to the number of observations within each cluster (100 clusters, each with 7 observations). Rodriguez and Goldman (1995) recommended the method as the most appealing option for more complicated models in their assessment of estimation procedures for multilevel models with binary responses, despite the intensive computation involved. However, Rodriguez and Goldman noted that no computer program for implementing the algorithm is publicly available. In this thesis, Zeger and Karim's algorithm is coded, evaluated and

implemented to study the distribution of learning opportunities. A small scale simulation study is carried out to assess the accuracy of the code and to examine how well the algorithm suits the present analysis, where the number of observations per cluster is larger than the total number of clusters (42 clusters, each with 100 observations on average).

The approach can be applied to policy research that seeks to understand how various policies and factors at the macro (e.g., nations, states, districts) and meso levels (e.g., schools and classrooms) of the educational system(s) operate and interact with one another, and whose design has relatively few self-selected macro units. For instance, the approach can be applied to study how the probability of employment of adults is related to one's educational attainment, social and ethnic background, literacy level, as well as state policies on continuing education and retraining programs, using the 1992 National Adult Literacy Study (Raudenbush, Kasim, Eamsukkawat & Miyazaki, 1996). It can be applied to panel data to study trends as well. An example, which will be an extension of the inquiry of this thesis, is to utilize the various waves of the TSAP in mathematics data (1990, 1992, 1996) to review the trends of the likelihood of offering algebra over time as a function of state reform initiatives.

1.4 Overview of the Study

Chapter 2 describes the background and significance of the substantive inquiry into learning opportunities as well as the development of a Gibbs sampler for Bayesian hierarchical models with binary outcomes. Chapter 3 outlines the Gibbs sampler technique developed by Zeger and Karim (1991), and presents the results of a simulation

study documenting the performance of the SAS\IML code written to implement the Gibbs sampler. Chapter 4 reports and discusses the findings on school-to-school and state-to-state variation in the likelihood of offering algebra, and the relationships between various school- and state-level factors and a school's decision to include algebra in its middle grades mathematics curriculum. Chapter 5 concludes the study and suggests future research needs.

CHAPTER 2

BACKGROUND AND SIGNIFICANCE

2.1 Introduction

This chapter describes the theoretical frameworks and reviews relevant literature for the substantive inquiry and methodological studies. It elaborates the research objectives and hypotheses stated in Chapter 1. It is contended that the inquiry can add to our understanding in how educational institutions function as distributors of learning opportunities, and the methodological work done in this thesis can be beneficial to policy research that addresses the multilevel structure of social and educational systems.

2.2 The Substantive Inquiry

2.2.1 Background

With the philosophy that access to algebra in middle grades is critical to a broad range of academic, and eventually occupational pursuits, and the mission to make algebra available to all middle grades students, projects like the Algebra Project (Jetter, 1993; Moses, 1994; Moses, Kamii, Swap, & Howard, 1989; Silva & Moses, 1990), and the Urban School Science and Mathematics Program (Archer, 1993) have made serious efforts to enhance mathematics and science education in the middle grades. Moses et al. and Archer postulated that, given the sequential nature of the mathematics curriculum, enrollment in algebra in middle grades is important in determining students' subsequent access to college preparatory mathematics in high schools. Empirical support for their

premise was provided by Stevenson, Schiller and Schneider's (1994) research on sequences of opportunities for learning mathematics from Grade 8 to Grade 10. They found, after controlling for achievement, coursework in eighth-grade algebra is associated with a greater likelihood of enrollment in advanced high school mathematics.

The mission "algebra for all" was initiated in part by the concern for equity in mathematics and science. A large number of inner-city, minority, and poor students are denied access to prerequisite courses in rigorous academic programs. Differential access to those courses is tied to race and class (Oakes, 1990, Raudenbush et al., 1996, Useem, 1992). The creators of the projects posed a challenge to the ability model, which decrees only mathematically inclined students should be given access to algebra, and its "institutional expressions" (Moses et al., p. 424)--curriculum tracking. They were convinced that, with appropriate pedagogy, capable teachers, high expectations of success, and strong support from school, parents and community, mastery of seventh- and eighth-grade algebra is within the reach of every student. MacIver and Epstein (1995) conducted an initial test of their belief that it is beneficial for public school students of all academic tracks to be given access to algebra in the middle grades, notwithstanding their past success in mathematics. Employing data from the base year of the National Educational Longitudinal Study of 1988 (NELS:88), MacIver and Epstein found that students who had taken eighth-grade algebra scored .3 to .5 standard deviations higher on a standardized mathematics test than those who did not, controlling for school characteristics, such as social and ethnic composition, and students' track level, past achievements in mathematics, and background variables. The increment in achievement,

moreover, was similar for students in different ability groups. Usiskin (1987) shared, to a great extent, the views of Moses et al. and Archer. Citing the schooling experiences of other countries and that of the University of Chicago School Mathematics Project, he contended that algebra can be learned by average students at the eighth-grade level.

The proponents of these mathematics projects stressed the critical role of school as a provider as well as a distributor of educational opportunity (MacIver & Epstein, 1995) in achieving equity as well as preparing the underrepresented non-Asian minorities for careers in science (Kamii, 1990; Oakes, 1990). As Porter (1991) succinctly stated, "Schools provide educational opportunity; they do not directly produce student learning" (p.33). Components that constitute opportunities to learn include exposure to content, pedagogy, instructional time, adequate institutional resources, and methods of assessment (Porter, 1991, 1994).

Sørenson and Hallinan's (1977) model for the process of learning highlights the impact schools can have through the instructional resources and learning environments they provide. This model was further elaborated by Hallinan (1996), who collaborated with Sørenson in its conceptualization. In the model, two individual attributes of the students--level of ability and effort--together with opportunities for learning provided by schools, constitute three primary determinants of student learning. The amount of learning is a function of these three interdependent elements. While the three determinants interact with one another in the learning process, the model, as Sørenson and Hallinan stressed, is not an additive one. A high level of ability and effort does not compensate for a lack of educational opportunity. Students cannot acquire the knowledge

from schooling if they are not exposed to the appropriate curriculum, no matter how able or diligent the students are. Opportunities to learn can enhance and the lack thereof can constrain a student's learning environment. As most low-income minority families cannot afford to supplement school experiences with learning opportunities in the private sector, such as hiring a tutor, the learning resources at school have a much larger impact on the academic achievements of disadvantaged students (Gordon, 1967, Oakes, 1990).

Another important concern of the projects is the immediate need to tackle the problem of the underparticipation of minorities in the fields of science, mathematics, and engineering. With the demographic change of increasing proportion of minorities and immigrants in the new labor force, Kamii (1990) argued that, in order to maintain the competitiveness and leadership role of the United States in the global economy, enticing the underrepresented groups into the "pipeline" leading to careers in science is particularly urgent.

The present investigation compares public schools across 42 states with respect to the opportunities they provide students to learn advanced mathematics in middle grades through offering eighth-grade algebra. It seeks to understand how the schooling systems in various states distribute to their students access to the critical gatekeeping algebra course. The course availability measure (Monk, 1994) used here is whether a school offers algebra for high school credits. The measure indexes the upper bounds of student learning in advanced mathematics set by middle and junior high schools (Gamoran, 1987), as the absence of the course precludes most possibilities to learn algebra in the eighth grade. Influences of various school- and state-level characteristics and policies on

this limit are considered. Past relevant studies, on which the formulation of research hypotheses in this study are based, are reviewed in the following section.

2.2.2 Relevant Literature

Two studies laid the groundwork for this dissertation: Raudenbush, Fotiu, et al.'s (1996) work on inequality of access to educational opportunity, and Mullis et al.'s (1994) report on effective schools in mathematics. Using the 1992 TSAP in mathematics of NAEP, Raudenbush et al. investigated social and ethnic inequality in access to resources for learning eighth-grade algebra. Of particular relevance here is the school resource--course offerings that they considered. Using the proportions of students classified according to ethnicity and level of parental education as predictors, the study modeled the probability of attending a school that offers high school algebra for eighth graders. The results indicates that the probability of attending a school with a rigorous mathematics program is positively associated with the level of parental education and that modest ethnicity gaps exist. In addition, there are substantial differences among states in the overall probability of students' access to algebra in grade eight.

Mullis et al. (1994) carried out a school effectiveness study pertaining to mathematics education using the 1992 NAEP 4th, 8th (not the TSAP in mathematics used in this study), and 12th grade mathematics achievement data from 1500 public and private schools. In the study, three series of analyses were carried out to classify the more effective versus the less effective schools, to compare the characteristics of those groups of schools, and to identify sets of variables associated with higher mathematics

achievement in the more effective schools.

In the first series of analyses, the researchers used the school mean mathematics proficiency scores to classify the schools participating in the study into three categories: highest-performing, medium-performing, and lowest-performing schools, and did a direct comparison of the contexts for learning mathematics, e.g., students' course enrollment, in the first and last groups of schools. In the second and third series of analyses, hierarchical linear modeling techniques were employed to identify factors associated with school effectiveness, controlling for home background of students and level of school SES.

The three series of analyses yielded congruent results. Their findings on school effectiveness as related to opportunities to learn, socioeconomic characteristics of students and schools, and school ethnic composition are reported here. With course taking as an opportunity-to-learn measure, the study found that the top one-third of schools had more students than their bottom one-third counterparts enrolled in advanced courses. For example, 27 versus 13 percent of ninth graders in the two different groups of schools had enrolled in algebra by eighth grade. Also, more economically disadvantaged, urban, and minority students were in the lower-performing schools. After adjusting for socioeconomic characteristics of students and schools, two of the eleven factors which significantly differentiated the two groups of the most and least effective schools for grade 8 were proportion of students enrolled in Algebra 1 and percentage of non-African Americans. The more effective schools had more students enrolled in Algebra 1 and fewer minority students (non-African Americans). The last series of analyses investigated predictors of mathematics achievement in more effective schools. In addition to the large

influences of socioeconomic factors, planning to enroll in or taking more advanced courses was a powerful predictor of higher mathematics achievement at grades 8 and 12. In sum, their study suggests that opportunities to learn, as measured by course enrollment, are likely to be linked to school achievement and are socially and racially stratified (Gamoran, 1987).

A few studies in addition to the two pieces of groundwork have implications for the formulation of research hypotheses of this dissertation. The results of Education in the Middle Grades: A National Survey of Practices and Trends (Becker, 1990) show substantial between-school variability in the offering of algebra. The survey found that out of a probability sample of 2,400 public schools that enrolled seventh-graders, 63% of the middle-grade schools offered algebra courses in their curriculums in grade 7, and, more typically, grade 8. More than one third of the schools in the sample did not offer algebra. The finding, together with Raudenbush, Fotiu, et al. 's (1996) report of substantial interstate variation in the probability of the offering of algebra, indicates that there is considerable variance at both the school and state level, and encourages an examination of its sources.

2.2.3 Sources of School-to-School Variation in the Availability of Eighth-Grade Algebra

At the school level, variability in the availability of the algebra course may reflect differences in school composition, urbanicity, and size, as conceptualized in Gamoran's (1987) model of opportunities for learning. Gamoran postulated that the location of a

school in an urban, suburban, or rural setting, and school composition may have influences on school offerings, and lead to between-school variation and stratification of learning opportunities. School offerings in turn may have an impact on students' achievement. Gamoran labeled and tested these indirect "setting effects" on achievement in his study. Oakes (1990), using data from the 1985-1986 National Survey of Science and Mathematics Education (NSSME), studied the percentages of 1,200 junior high schools of different SES (high poverty to high wealth) and Caucasian student populations (different categories of different proportions) that had one or more sections of advanced mathematics, i.e., eighth-grade algebra and ninth-grade geometry. An analysis of variance revealed no significant differences in the percentages of schools that offered accelerated mathematics classes among the SES or racial composition categories. She, however, suggested that meaningful differences may exist. Specifically, low-SES and predominantly minority schools are less likely to offer accelerated mathematics classes. In another analysis, Oakes examined advanced mathematics sections per 100 students and found that high-SES or all-white schools have significantly more of those advanced sections. Useem (1992) reported that district parental education levels were related to the striking district-by-district variability of student enrollment in eighth-grade algebra, which ranged from 13% to 60%.

The present study cross-validates Oakes' findings of non-significant SES and racial composition effects on course offering and evaluates the hypothesis that racial and social stratification exist in the offering of algebra. Given the other related results and that of Raudenbush, Fotiu, et al. (1996), it is expected that schools with fewer economically

disadvantaged students and African and Hispanic minority students are more likely to have rigorous mathematics programs.

Monk and Haller (1993) analyzed the 1980 High School and Beyond (HSB) survey data with 1032 public and private schools in the U.S. and found a positive significant interaction effect of urban setting and school size on the total number of academic course credits. Monk and Haller argued as urban high schools face a less restrictive local teacher labor market, which lifts the human resources constraint on curriculum development, diversity in academic course offerings is favored there. Therefore, all else being equal, the likelihood of offering algebra is hypothesized to be higher in urban versus rural and, similarly, suburban versus rural schools.

Another source of variability among schools in the likelihood of the offering of algebra may originate from a structural factor of schools, that of size. Monk and Haller (1993) contended that greater school size increases efficiency through its potential to generate economies, and their study showed a positive relationship between the enrollment of a high school's graduating class and how many different course credits were offered. Furthermore, Useem (1992) in her study on variability in advanced mathematics course placements across 26 districts, reported that in one K-8 system only some of the larger elementary schools offered algebra. The two studies suggest that greater school size is associated with the offering of a comprehensive and specialized curriculum, as contended by Lee and Smith (1995). It is therefore hypothesized that schools with larger grade 8 enrollments are more likely to offer algebra, net of the effects of school racial and social composition and size.

2.2.4 Sources of State-to-State Variation in the Availability of Eighth-Grade Algebra

Two possible sources of state variation in the availability of algebra are state investment efforts in public education and poverty levels. As financial support is related to the quality of resource inputs, such as the qualifications of the mathematic teachers and how updated the instructional materials are, the amount being invested in the educational systems is likely to have an impact on the distribution of scarce educational resources. This is particularly important for schools in economically disadvantaged settings where there are typically fewer educational resources and less qualified mathematics teachers (Archer, 1991; Tate, 1994). Indeed, Tate posited that without fiscal equity for those schools, the chances for success of implementing the curricular reforms recommended by the well-known Curriculum and Evaluation Standards for School Mathematics (National Council of Teachers of Mathematics (NCTM), 1989), are slim. His views echo those of the proponents of systemic reforms (e.g., O'Day & Smith 1993).

Given that state educational spending accounts for nearly half of total K-12 public school revenues (National Center for Education Statistics, 1994), the present inquiry tests the exploratory hypothesis that different levels of financial resources at the state level are associated with different likelihoods of offering of algebra. How much a state spends on education may also influence the availability of the course through its impacts on the relationships between school and ethnicity composition and the likelihood of the course. In other words, state expenditures may interact with the effects of student social and ethnic composition, if any, in influencing a school's decision to include algebra in its mathematics curriculum. The exploratory hypothesis aims at investigating the indirect

effects of expenditures on opportunities to learn. Intervening processes, such as the possibility that increased expenditures may help prepare and recruit better qualified teachers and thus may supply the instructional staff necessary for the offering and teaching of advanced courses, are not tested here. Another exploratory hypothesis pertaining to the poverty level of the state is that schools in states with higher poverty rates, on average, are less likely to offer algebra.

2.2.5 Significance of the Substantive Inquiry

While built upon the existing research, the inquiry of this thesis differs from, and thus, adds to the literature on learning opportunities in several ways. First, unlike most studies, except the one by Raudenbush, Fotiu, et al. (1996), it adopts a multilevel framework and incorporates the clustering effects of states. It explicitly models the state memberships shared by schools, an unprecedented opportunity presented by the TSAP in mathematics. The framework allows one to study interstate and interschool variation appropriately (see Raudenbush, 1988) and how they can be explained by state characteristics and policies.

Whereas both Raudenbush, Fotiu, et al.'s (1996) and the present study investigate access to algebra, they adopt different approaches. The former focused on person-to-person variation and investigated the probability of placement of students in schools which offered algebra as a function of the students' background characteristics. The present work surveys what school- and state-level factors may affect a public school's decision to offer high school algebra. The study focuses on one of the best "school-

controlled predictors of student achievement" (Porter, 1994, p.427)--content of instruction.

Another major divergence from the previous research on course availability such as Oakes' (1990) studies is the use of more comprehensive models in the present study. The models assessed contain various school compositional, setting, and structural predictor variables. It allows one to investigate the influences of individual variables net those of the others. For instance, one may be interested in assessing the ethnicity bias while holding constant the influences of the structural characteristic of a school. The multilevel and better specified models in the study allow a better understanding of the school decision-making process in delivering learning opportunities to their students, and how the process may be related to state financial investment in education.

Furthermore, as state revenue makes up a significant portion of the total school revenue, states can exert influences on how those monies are transformed into educational opportunities. One way to do so, as argued by Monk (1994), is through the explicit statements of the basis of the financial entitlement. A state, for instance, might enumerate the sort of educational opportunities, or the "intended curriculum" (Pelgrum et al., 1995), a district should provide and collect information on how well the district meets the specified details. It is believed that the measure of course availability, operationalized as the likelihood of the offering of algebra, can supply some of the relevant information and act as a production and curriculum indicator (Oakes, 1989; Pelgrum et al., 1995; Porter, 1991).

2.3 The Methodological Work

2.3.1 Motivations

The substantive inquiry has motivated the development and evaluation of a fully Bayesian analytical framework which accommodates 1) the binary nature of the outcome variable of the offering of the algebra course, 2) the hierarchical or nested design of the TSAP in mathematics--schools nested within states and territories, and 3) the need to incorporate the uncertainty about the variances at the state level. By representing the state-level variance as the investigator's uncertainty about the processes that produce it, the hierarchical framework enables us to study state-level heterogeneity, the relationships between state-level predictors and outcomes (Raudenbush et al. 1995), and not to treat the states as fixed strata. To illustrate the framework and to establish notation, I use an over-simplified model for the availability of high-school algebra, with two predictor variables: percentages of minorities (Hispanic and African American students) per school and state educational expenditures per student. The fully Bayesian estimation approach, together with the ML and AML\PQL strategies, is then outlined.

2.3.2 An Illustrative Example

Let y_{ij} be an indicator outcome variable which takes on a value of 1 if high-school algebra is offered, and 0 otherwise for school i in state j, where $i = 1$ to n_j and $j = 1$ to J. The goal of the analysis is to explore the correlates of the probability, p_{ij}, that a particular school offers eighth-grade algebra. At the school level, it is of interest to determine if higher percentage of minorities in a school $HiMin_{ij}$ is associated with less likelihood of

the algebra course being offered, or if the learning opportunities for the gatekeeping course to advanced high school mathematics are racially stratified (Gamoran, 1987). $HiMin_{ij}$ is an indicator variable and is coded 1 for schools with more than 50% Hispanic and African American students, and 0 otherwise. A discussion of the coding scheme of this variable is given in Chapter 4. To constrain the probability to lie within a [0,1] interval, p_{ij} is transformed to η_{ij} using the logit link function. The transformed outcome is now the log-odds of the offering of algebra. The model is expressed as:

$$\log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \eta_{ij} = \beta_{0j} + \beta_{1j} HiMin_{ij} , \qquad \textbf{(2.1)}$$

or in vector form,

$$\eta_{ij} = (1 \quad HiMin)_{ij} \begin{bmatrix} \beta_{0j} \\ \beta_{1j} \end{bmatrix} . \qquad \textbf{(2.2)}$$

Each coefficient defined in the school-level equation is modeled as an outcome at the state level. The intercept of the school-level model, β_{0j}, which is the average log-odds of the offering of algebra for schools which are located in state j and have a relatively low percentage of Hispanics and African American students, is estimated to see whether it is related to fiscal equity. The intercept is modeled as a function of state educational expenditure per student $StateExp_j$, plus a random state effect according to the model

$$\beta_{0j} = \gamma_{00} + \gamma_{01} StateExp_j + u_{0j} , \qquad \textbf{(2.3)}$$

where $u_{0j} \sim N(0, \tau)$. In addition, whether or not educational expenditure may be related to

possible racial stratification in learning opportunities, as represented by β_{1j}, is assessed. The following model expresses the relationship:

$$\beta_{1j} = \gamma_{10} + \gamma_{11} StateExp_j \ . \qquad \textbf{(2.4)}$$

In matrix notation, the combined state-level model can be expressed as

$$\begin{bmatrix} \beta_{0j} \\ \beta_{1j} \end{bmatrix} = \begin{bmatrix} 1 & StateExp_j & 0 & 0 \\ 0 & 0 & 1 & StateExp_j \end{bmatrix} \begin{bmatrix} \gamma_{00} \\ \gamma_{01} \\ \gamma_{10} \\ \gamma_{11} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} u_{0j} \end{bmatrix}, \quad \textbf{(2.5)}$$

and the combined school- and state-level model can be stated as

$$\eta_{ij} = (1 \ StateExp_j \ HiMin_{ij} \ HiMin_{ij} * StateExp_j) \begin{bmatrix} \gamma_{00} \\ \gamma_{01} \\ \gamma_{10} \\ \gamma_{11} \end{bmatrix} + \quad \textbf{(2.6)}$$

$$[1] \begin{bmatrix} u_{0j} \end{bmatrix},$$

which can be formulated as a more general two-level mixed model with a logit-normal hierarchy (Searle, 1991),

$$\eta_{ij} = \boldsymbol{x}_{ij}^{\mathrm{T}} \gamma + \boldsymbol{z}_{ij}^{\mathrm{T}} \boldsymbol{u}_j \ , \qquad \textbf{(2.7)}$$

where

$\boldsymbol{x}_{ij}$ is a p x 1 vector of predictors associated with the log-odds of the algebra course being offered; for model **(2.6)**, $p = 4$;

γ is a p x 1 vector of regression coefficients for fixed effects;

z_{ij}, a subset of x_{ij}, is a r x 1 vector of predictors associated with r random effect(s); for model **(2.6)**, $r = 1$; and

u_j is a r x 1 vector of random effects and it is normally distributed with mean of zero and variance of T. For **(2.6)**, the variance is a scalar τ.

The parameters of interest in model **(2.7)** are γ, u_j, and T. The marginal likelihood for the parameters γ and T is proportional to

$$l(\gamma,T) \propto \prod_{j=1}^{J} \int \prod_{i=1}^{n_j} f(y_{ij} \mid u_j) g(u_j) du_j , \qquad (2.8)$$

where

$$f(y_{ij}|u_j) = \left(\frac{1}{1 + \exp^{-\eta_{ij}}} \right)^{y_{ij}} \left(\frac{\exp^{-\eta_{ij}}}{1 + \exp^{-\eta_{ij}}} \right)^{1-y_{ij}} , \qquad (2.9)$$

and

$$g(u_j) \propto |T|^{-1/2} \exp[-\frac{1}{2} u_j^T T^{-1} u_j] du_j . \qquad (2.10)$$

Model **(2.7)** is a special case of the generalized linear model (GLM) (McCullagh & Nelder, 1989, Nelder & Wedderburn, 1972) with random effects. GLM has unified models with outcomes having a variety of different scales. Some examples are the linear regression models for approximately Gaussian data (e.g., SAT scores), logit models for dichotomous data (e.g., offering of algebra), log-linear models for count data (e.g., days of absence from school), and proportional hazards models for survival time data (e.g.,

years taken to attain a doctoral degree). The unification has brought with it a common theoretical framework and estimation procedure (see McCullagh & Nelder, 1989). Model **(2.7)** is an extension of GLM with an incorporation of random terms $\boldsymbol{u}_j$ to handle clustered data. The conditional distribution of y_{ij} given $\boldsymbol{u}_j$ assumes an exponential family distribution of the form

$$f(y_{ij}|\boldsymbol{u}_j) = exp([y_{ij}\theta_{ij} - a(\theta_{ij}) + b(y_{ij})]/\varphi) , \qquad \textbf{(2.11)}$$

where

θ_{ij} is the natural parameter,

φ is the dispersion parameter, and

$a(\cdot)$ and $b(\cdot)$ are specific functions corresponding to the type of exponential family.

The specific parameters of the conditional distribution are:

$$\begin{aligned} \theta_{ij} &= \eta_{ij} , \\ a(\theta_{ij}) &= \log(1 + \exp^{\eta_{ij}}) , \end{aligned} \qquad \textbf{(2.12)}$$

and

$$\begin{aligned} b(y_{ij}) &= 0 , \\ \varphi &= 1 . \end{aligned} \qquad \textbf{(2.13)}$$

Modeling the natural parameter, θ_{ij}, directly as in **(2.12)** makes the link, η_{ij}, a canonical one (McCullagh and Nelder, 1989). The conditional moments $p_{ij} = E(y_{ij}|\boldsymbol{u}_j) = a'(\theta_{ij})$ and v_{ij}

$= var\,(y_{ij}|\boldsymbol{u}_j) = a''(\theta_{ij})\varphi$ are:

$$a'(\theta_{ij}) = \frac{1}{1 + \exp^{-\eta_{ij}}} = p_{ij} ,$$
$$a''(\theta_{ij})\varphi = \frac{\exp^{-\eta_{ij}}}{(1 + \exp^{-\eta_{ij}})^2} = p_{ij}(1 - p_{ij}) . \qquad \textbf{(2.14)}$$

Three main estimation approaches for the model are 1) the ML approach (e.g., Anderson & Aitkin, 1985; Fahrmeir & Tutz, 1994; Gibbons & Hedeker, 1994; Hedeker & Gibbons, in press; Karim, 1991), 2) the AML and the closely related PQL approach (e.g., Breslow & Clayton, 1993; Goldstein, 1991; McGilchrist, 1994; Schall, 1991; Stiratelli, Laird, & Ward, 1984; Raudenbush, 1993; Wolfinger & O'Connell, 1993), and 3) the Bayes approach (e.g., Dellaportas & Smith, 1993; Zeger & Karim, 1991).

2.3.3 The Maximum Likelihood Approach

The ML approach maximizes the marginal likelihood **(2.8)** or its logarithm with respect to the fixed effects $\boldsymbol{\gamma}$ and the variance components $\mathbf{T}$. Let $y_j = (y_{1j},...,y_{nj})^T$ and $\boldsymbol{\phi} = (\boldsymbol{\gamma}, \mathbf{T})$, the approach maximizes

$$log\ l(\boldsymbol{\phi}) = \sum_{j=1}^{J} log\ f(y_j|\boldsymbol{\phi}) . \qquad \textbf{(2.15)}$$

However, the marginal likelihood **(2.8)** does not have an analytic solution. Thus, numerical or Monte Carlo integration methods are needed for approximations. Two

maximization strategies, direct and indirect, are available for parameter estimation (Fahrmeir & Tutz, 1994). Whereas the direct strategy evaluates the marginal likelihood **(2.8)** and its partial derivatives via numerical or Monte Carlo methods in parameter estimation, the indirect one does not.

Fahrmeir and Tutz (1994) provided the algorithmic details for the direct maximization strategy. The procedure starts with a reparameterization of the random effects $\boldsymbol{u}_j$ in the conditional mean $E(\boldsymbol{y}_j|\boldsymbol{u}_j)$ as given in **(2.14)**:

$$\boldsymbol{u}_j = \mathrm{T}^{1/2}\boldsymbol{a}_j , \qquad \textbf{(2.16)}$$

where $\mathrm{T}^{1/2}$ is the left Cholesky factor of T and $\boldsymbol{a}_j$ is a standardized random vector with zero mean and the identity matrix as covariance matrix. The two-level mixed model **(2.7)** now becomes

$$\eta_{ij} = \boldsymbol{x}_{ij}^{\mathrm{T}}\gamma + \boldsymbol{z}_{ij}^{\mathrm{T}}\mathrm{T}^{1/2}\boldsymbol{a}_j , \qquad \textbf{(2.17)}$$

and the marginal log-likelihood now becomes $l(\gamma,\Omega)$, where $\Omega = \mathrm{vec}(\mathrm{T}^{1/2})$,

$$l(\gamma,\Omega) \propto \prod_{j=1}^{J} \int \prod_{i=1}^{n_j} f(y_{ij}|\boldsymbol{a}_j)g(\boldsymbol{a}_j)d\boldsymbol{a}_j , \qquad \textbf{(2.18)}$$

Specifying g and maximizing $l(\gamma,\Omega)$ with respect to the various parameters yield maximum likelihood estimates for γ and Ω, and subsequently T. The maximization can be accomplished by an iterative procedure such as Newton-Raphson or Fisher scoring method, and the evaluation of the integral in **(2.18)** can be carried out by the Gauss-Hermite quadrature technique when g is normal. The quadrature technique approximates

an integral by summing a certain number of quadrature points d for each dimension of the integration. If the number of quadrature points d is large enough, the approximation can be made increasingly precise and the approach could yield maximum likelihood estimates that have the properties of consistency, asymptotic normality, and efficiency. Using consistent estimators $\hat{\gamma}$ and $\hat{\tau}$, empirical Bayesian point estimator of $\boldsymbol{u}_j$ can be computed, which allows the recovery of estimates of $\boldsymbol{u}_j$. The algorithm allows likelihood-ratio tests for comparisons of nested models differing in variance and covariance components in addition to fixed effects.

A program called MIXOR (Hedeker and Gibbons, in press), which employs a direct maximization strategy called marginal maximum likelihood (MML) for estimating random-effects logit regression models, is available. The algorithm is very similar to the one of Fahrmeir and Tutz (1994). The models are basically the same as model **(2.7)**. In their derivation, the models differ from **(2.7)** in their introduction of an unobservable or latent continuous variable which "trigger[s] the discrete response" (Cramer, 1991, p.11). A threshold value is assumed and a discrete response occurs if the value of the latent continuous variable exceeds the threshold value. An example of a latent continuous variable is a household's desire to own a car and there is a threshold value or certain critical level of desire beyond which ownership may result (Cramer, 1991).

Karim (1991) developed an indirect maximization strategy to estimate the various parameters using a Monte Carlo implementation of the EM algorithm (MCEM) (Wei & Tanner, 1990, Dempster et al., 1977). In general, the EM algorithm is an iterative procedure for finding maximum likelihood estimates in the presence of unobserved or

missing data in probability models. The idea is to augment the observed data with latent data in order to replace one complicated maximization by an iterative series of simple maximizations (Tanner, 1990). Let $\boldsymbol{y} = (y_1, ..., y_j)^T$ and $\boldsymbol{u} = (u_1, ..., u_j)^T$, for model **(2.7)**, $\boldsymbol{y}$ is the observed or incomplete data, $\boldsymbol{u}$ is the latent data, and $(\boldsymbol{u}, \boldsymbol{y})$ is considered to be the complete data. The complete data log-likelihood is given by $\log f(\boldsymbol{u}, \boldsymbol{y}|\phi)$ and the conditional predictive distribution of the unobserved or missing data is $f(\boldsymbol{u}|\boldsymbol{y},\phi^{(i)})$, where $\phi^{(i)}$ is a current guess of the parameters. The iterative series of the simplified maximizations consists of two steps: the E-step (for expectation) and the M-step (for maximization). In the E-step the expected value of the complete data log-likelihood with respect to the conditional predictive distribution is determined, i.e.,

$$Q(\phi|\phi^{(i)}) = E(\log f(\boldsymbol{u},\boldsymbol{y}|\phi)|\boldsymbol{y},\phi^{(i)}) , \qquad \textbf{(2.19)}$$

which can be obtained by evaluating

$$Q(\phi|\phi^{(i)}) = \int \log f(\boldsymbol{u},\boldsymbol{y}|\phi) f(\boldsymbol{u}|\boldsymbol{y},\phi^{(i)}) d\boldsymbol{u} . \qquad \textbf{(2.20)}$$

In the M-step, $Q(\phi|\phi^{(i)})$ is maximized by finding the solutions to

$$\frac{\delta}{\delta\phi} Q(\phi,\phi^{(i)}) = 0 , \qquad \textbf{(2.21)}$$

and $\phi^{(i+1)}$ is thus determined. This completes one cycle and the iterative series continues until convergence. Iterations between E- and M-steps lead ultimately to the maximum

likelihood values of the parameters under mild regularity conditions (Dempster et al., 1977; Wu, 1983). The algorithm does not yield any variances and covariances of the estimators, but they can be obtained from the Hessian of the score functions of the various parameters.

For model **(2.7)**, however, no closed form exists for the E-step in **(2.20)**. Karim (1991) adopted a Monte Carlo implementation of the E-step (Wei & Tanner, 1990) as a result. The general scheme of Monte Carlo integration involves generation of random variates, which adds another step to the EM iterative series. The modified EM series begins with the generation of a sample of latent data, $\boldsymbol{u}^{(1)}$, ..., $\boldsymbol{u}^{(m)}$, where m = 1 to M from the current approximation to the conditional predictive distribution $f(\boldsymbol{u}\,|\boldsymbol{y},\phi^{(i)})$. The expected value of the complete data log-likelihood is then obtained as a mixture of the complete data log-likelihood over the generated latent data, i.e.,

$$Q_{(i+1)}(\phi,\phi^{(i)}) \cong \frac{1}{M}\sum_{m=1}^{M} \log f(y,u^{(m)}|\phi) \ . \qquad \textbf{(2.22)}$$

Then the M-step maximizes $Q_{(i+1)}(\phi|\phi^{(i)})$ and the new maximizer is used to update the conditional predictive distribution.

Karim (1991) used an indirect sampling method called importance sampling (Ripley, 1981) to generate latent data from $f(\boldsymbol{u}\,|\boldsymbol{y},\phi^{(i)})$. He approximated the posterior with a Gaussian distribution with mean and variance equal to the posterior mode and the posterior curvatures of the conditional predictive distribution. Then random variates were sampled from the approximated Gaussian distribution. To adjust for the discrepancy

between the true posterior and the approximating distribution, each simulated draw was assigned a weight. Let $g(\cdot)$ be the Gaussian function, $f(\cdot)$ be the conditional predictive function, and let $g(\hat{\boldsymbol{u}})$ be a draw from $g(\boldsymbol{u})$, the adjustment weight was computed as $f(\hat{\boldsymbol{u}})/g(\hat{\boldsymbol{u}})$.

Karim (1991) noted that with MCEM, the estimates fluctuate around the true maximum even after convergence. He recommended various tests including visual inspections of the traces of the estimates and assessment of variability in consecutive iterations. He implemented a simulation study and showed that the algorithm yielded reasonable inferences. No software is publicly available for this algorithm.

2.3.4 The Approximate Maximum Likelihood Approach\Penalized Quasi-likelihood Approach

Another main approach to parameter estimation for model **(2.7)** is the AML (e.g., Goldstein, 1991; McGilchrist, 1994; Schall, 1991; Stiratelli et al., 1984; Wolfinger & O'Connell, 1993), or the closely related PQL approach (e.g., Breslow & Clayton, 1993; Raudenbush, 1993). The essence of the approach is to derive an approximation to the marginal likelihood **(2.8)** and its partial derivatives, which allows repeated application of normal theory (Kuk, 1995). The approximation can be achieved by the linearization of the model (e.g., Goldstein, 1991), or by the use of penalized quasi-likelihood (e.g., Breslow & Clayton, 1993).

The linearization can be implemented by a Taylor series expansion of p_{ij} in **(2.14)** around the i^{th} current guess of $\boldsymbol{\gamma} = \boldsymbol{\gamma}^{(i)}$ and $\boldsymbol{u}_j = \boldsymbol{u}_j^{(i)}$. p_{ij} is approximated in the region of a

point in $\eta_{ij}^{(i)}$. The procedure, which can be implemented by the Newton-Raphson algorithm, yields a set of linearized dependent variables y_{ij}*'s and weights w_{ij}'s for each data point (McCullagh & Nelder, 1989; Raudenbush, 1993; Schall, 1991), where

$$y_{ij}^{*} = \eta_{ij}^{(i)} + \left(\frac{\partial p_{ij}^{(i)}}{\partial \eta_{ij}^{(i)}}\right)^{-1}(y_{ij} - p_{ij}^{(i)}), \qquad \textbf{(2.23)}$$

with

$$\begin{aligned} \eta_{ij}^{(i)} &= \boldsymbol{x}_{ij}^{T}\gamma^{(i)} + \boldsymbol{z}_{ij}^{T}\boldsymbol{u}_{j}^{(i)} , \\ p_{ij}^{(i)} &= (1 + \exp^{-\eta_{ij}^{(i)}})^{-1} , \\ \frac{\partial p_{ij}^{(i)}}{\partial \eta_{ij}^{(i)}} &= w_{ij}^{(i)} = p_{ij}^{(i)}(1 - p_{ij}^{(i)}) , \end{aligned} \qquad \textbf{(2.24)}$$

Lindstrom and Bates (1990) pointed out that this Taylor series expansion can be conceptualized as a step for creating psuedo-data.

Alternatively, the approximation can be achieved by replacing $f(y_{ij}|\boldsymbol{u}_j)$ in **(2.9)** by the exponential of the quasi-likelihood that denotes the deviance measure of fit (Breslow & Clayton, 1993; McCullagh & Nelder, 1989), which depends only on the first two moments of the model. Laplace's method for integral approximation is then applied to the marginal likelihood with the substituted quasi-likelihood. The approximation then allows the use of normal theory to estimate the parameters.

Iterative algorithms such as generalized least squares (e.g., Goldstein, 1991), Fisher scoring (e.g., Breslow & Clayton, 1993), and EM-type algorithm (e.g.,

Raudenbush, 1993) can be applied to the approximated likelihood to obtain approximate maximum likelihood estimates. The EM-type algorithm is related to MCEM in that it is obtained when M in **(2.22)** is equal to one (Wei & Tanner, 1990). The iterative algorithm consists of forming and maximizing the function $\log f(\boldsymbol{u}_j=\hat{\boldsymbol{u}}_j,\boldsymbol{y}^*|\ \phi)$ over ϕ, where $\hat{\boldsymbol{u}}_j$ is the posterior mode of $f(\boldsymbol{u}_j|\boldsymbol{y},\phi^{(i)})$, to obtain $\phi^{(i+1)}$ and update $\hat{\boldsymbol{u}}_j$. This EM-type algorithm which relies on posterior modal analysis, coupled with the Newton-Raphson algorithm used for creating psuedo-data, are employed by Stiratelli et al. (1984) and Raudenbush (1993) to estimate model **(2.7)**.

The Taylor expansion for the approximation can be either about $\boldsymbol{u}_j = \boldsymbol{u}_j^{(i)}$ or $\boldsymbol{u}_j = 0$. The former type of model corresponds to the PQL model and the latter marginal quasi-likelihood (MQL) model in Breslow and Clayton's (1993) classification of models with and without the incorporation of the random effects terms $\boldsymbol{z}_{ij}^T\boldsymbol{u}_j$ in the linear predictor. Goldstein and Rabash (1996) showed that the PQL estimates with a second order Taylor expansion are less biased than those of the first order MQL.

An option to estimate restricted approximate maximum likelihood estimates for degree-of-freedom adjustment (Breslow & Clayton, 1993) is also available in this approach. In cases where the outcomes are normal, such adjustment can alleviate the downward bias of the maximum likelihood estimates of T. To implement the estimation, Stiratelli et al. (1984) and Raudenbush (1993) handled γ as random effects, whose variances tend to infinity, in the EM-type routine in their algorithms.

Several software packages for the approximate maximum and penalized quasi-likelihood estimation are available. HLM2/3 (Bryk, Raudenbush, & Congdon, 1996),

ML3 (Prosser, Rasbash, & Goldstein, 1991), and VARCL (Longford, 1988) are some examples. Several simulation studies were done to assess the estimators of this approach. Rodriguez and Goldman (1995) and Breslow and Clayton (1993) found that the MQL estimators display considerable bias, especially with reference to T and when the binomial denominators are small. Kuk (1995) and Goldstein and Rabash (1996) applied iterative boostrap bias correction and Kuk showed that the procedure yields asymptotically consistent and unbiased parameter estimates.

2.3.5 The Fully Bayesian Approach

The third major approach to parameter estimation is the fully Bayesian approach. Unlike the previous two approaches, which estimate T and/or γ as fixed parameters, Bayesian inference treats all parameters as random quantities with prior distributions. Prior distributions have two basic interpretations, the population and the state of knowledge (Gelman, Carlin, Stern & Rubin, 1995). In the first interpretation, "the prior distribution represents a population of possible parameter values, from which the [parameter] ... of current interest has been drawn" (Gelman et al., 1995). In the second interpretation, which is more subjective, "the guiding principle is that we must express our knowledge (and uncertainty) about the [parameter] ... as if its value could be thought of as a random realization from the prior distribution" (Gelman et al., 1995). The main inferential goal of Bayesian analysis can be considered to be updating the prior knowledge of the random quantities in light of the observations. The analysis requires computation of the joint posterior distribution, $p(\gamma,T,u|y)$, which is proportional to the

product of the likelihood function for the parameters and their prior distributions:

$$p(\gamma,\mathrm{T},\boldsymbol{u}\,|y) \;=\; k^{-1}\prod_{j=1}^{J}\prod_{i=1}^{n_j} p(y_{ij}|\gamma,\boldsymbol{u}_j)p(\boldsymbol{u}_j|\mathrm{T})p(\gamma,\mathrm{T}), \qquad \textbf{(2.25)}$$

where

$$k \;=\; \int\prod_{j=1}^{J}\int\prod_{i=1}^{n_j} p(y_{ij}|\gamma,\boldsymbol{u}_j)p(\boldsymbol{u}_j|\mathrm{T})p(\gamma\ \ \mathrm{T})\partial\boldsymbol{u}_j\partial\gamma\,\partial\mathrm{T}\ . \qquad \textbf{(2.26)}$$

In the joint distribution **(2.25)**, the joint prior for T and γ is $p(\mathrm{T},\gamma)$, and the prior for $\boldsymbol{u}_j$ is the product of the J distributions $p(\boldsymbol{u}_j\,|\mathrm{T})$. Through the likelihood function for the parameters T, γ, and $\boldsymbol{u}$, which is

$$l(\gamma,\mathrm{T},\boldsymbol{u}) \;=\; \prod_{j=1}^{J}\prod_{i=1}^{n_j} p(y_{ij}|\gamma,\boldsymbol{u}_j)p(\boldsymbol{u}_j|\mathrm{T})\ , \qquad \textbf{(2.27)}$$

the prior distributions are modified and updated in Bayesian posterior analysis (Box & Tiao, 1973). When inferences on single parameters are required, the joint posterior distribution is to be integrated with respect to the other or what is called nuisance parameters (Gelfand, Hills, Racine-Poon & Smith, 1991) to obtain the marginal posteriors $p(\gamma\,|y)$, $p(\mathrm{T}|y)$ and $p(\boldsymbol{u}\,|y)$,

$$p(\gamma\,|y)=\int\!\!\int p(\gamma,\mathrm{T},\boldsymbol{u}\,|y)\partial\boldsymbol{u}\,\partial\mathrm{T}\ , \qquad \textbf{(2.28)}$$

$$p(\mathrm{T}\,|y)=\int\!\!\int p(\gamma,\mathrm{T},\boldsymbol{u}\,|y)\partial\boldsymbol{u}\,\partial\gamma\ , \qquad \textbf{(2.29)}$$

and

$$p(\boldsymbol{u}|y) = \int\int p(\gamma, T, \boldsymbol{u}|y)\partial\gamma\,\partial T \quad . \qquad \textbf{(2.30)}$$

Characteristics of the joint and marginal posterior characteristics such as means, modes, and variances are legitimate for Bayesian inference. The integrals above do not have analytic solutions. Thus Bayesian inference requires numerical evaluation. The task may become formidable when the model increases in complexity. An alternative to avoid the difficulty is to employ the Gibbs sampling technique (Gemen & Gemen, 1984; Gelfand & Smith, 1990) to simulate the posteriors, therefore leading to draws of the parameters. The next chapter describes the application of the Gibbs sampler to random effects generalized linear models.

A simulation study done by Zeger and Karim (1991) showed that the algorithm yielded reasonable inferences in finite sample cases in which the number of clusters was large relative to the number of observations within each cluster (e.g., 100 clusters, each with 7 observations). When compared to AML or PQL, the algorithm yielded less biased estimates. No computer program for implementing the algorithm developed and tested by Zeger and Karim is publicly available.

How the three estimation approaches compare and contrast with one another in their inferential goals and strategies are summarized in Table 2.1. In the AML/PQL and ML approaches, the major objective is to maximize or find the maximizer of the likelihood functions and the posterior distributions. In the Bayesian approach, the main goal is to seek the marginal posteriors of the various parameters (Tanner, 1993). Unlike

ML and Bayes, data affect inference via the quasi-likelihood function in AML/PQL. Finally, $\mathbf{T}$ is treated as random only in the Bayesian approach. The advantages offered by the distinguishing features are given in the next section.

Table 2.1: Defining Features for the Three Major Approaches

Features/ Approach	AML\PQL	ML	Bayes
Objectives	Approximate likelihood **(2.8)** and its partial derivatives and maximize the approximated likelihood with respect to γ and T, then predict $\boldsymbol{u}_j$, given $\hat{\gamma}$ and $\hat{\mathrm{T}}$	Maximize the marginal likelihood **(2.8)** with respect to γ and T, then predict $\boldsymbol{u}_j$, given $\hat{\gamma}$ and $\hat{\mathrm{T}}$	Find the marginal posterior distributions of the parameters: $p(\gamma \mid y)$, $p(\mathrm{T} \mid y)$ and $p(\boldsymbol{u}_j \mid y)$
Data affect inference via the likelihood	No	Yes	Yes
All parameters are treated as random quantities	No	No	Yes

2.3.6 Appropriateness of the Fully Bayesian approach

Among the three main approaches, the Bayesian approach is deemed more appropriate for the inquiry studying the distribution of learning opportunities for the following reasons. First, relative to AML\PQL, Bayes estimates of the level-2 variances are less biased (Breslow & Clayton, 1993; Rodriguez & Goldman, 1995). This is because in the Bayesian approach, the data affect inference via the likelihood instead of the approximate or quasi-likelihood. Second, relative to ML and AML\PQL, Bayes inferences about any parameter fully take into account uncertainty about other parameters in the same model. In the AML\PQL procedures for restricted approximate maximum likelihood estimation, for instance, $\mathbf{T}$ is treated as fixed at its point estimates. Estimates of all regression coefficients and their standard errors are then conditioned on those point estimates. This poses two problems. First, Zeger and Karim (1991) pointed out that the estimates of the fixed effects γ and the variance $\mathbf{T}$ are asymptotically correlated and inferences about the fixed effects have to incorporate the uncertainty in the estimate of $\mathbf{T}$. In addition, as there are only 42 states in the data, the variance and co-variance components are likely to be estimated with moderate or poor precision. By treating $\mathbf{T}$ as random in the fully Bayesian approach, uncertainty arising from the state-to-state heterogeneity can be incorporated into the estimates of the regression coefficients. The incorporation of uncertainty is particularly important with regard to inferences about relationships between state-level predictors and outcomes, which is of great importance in this study. Lastly, the Bayesian estimation provides researchers with richer inferential information by outputting the entire posterior distributions of estimates, rather than point

and interval estimates alone. The Bayesian estimation approach is useful here, given the modest number of states and the posterior distribution of the level-2 variance is likely to be skewed. Inferences based on the variance of the estimates alone can be misleading.

2.3.7 Usefulness of the Bayesian Approach to Policy Research

The approach has much to offer to policy research that seeks to understand how various policies and factors at the macro (e.g., nations, states, districts) and meso levels (e.g., schools and classrooms) of the educational system(s) operate and interact with one another, and whose design has a relatively small number of self-selected macro units. For instance, it can be applied to study how the probability of employment of adults is related to one's educational attainment, social and ethnic background, literacy level, as well as state policies on continued education and retraining programs, using the 1992 National Adult Literacy Study (Raudenbush et al., 1996). It can be applied to panel data studying trends as well. An example, which will be an extension of the inquiry of this thesis, is to utilize the various waves of the TSAP in mathematics data (1990, 1992, 1996) to study the trends of the likelihood of offering algebra over time as a function of state reform initiatives.

CHAPTER 3

BAYESIAN ESTIMATION OF GENERALIZED LINEAR MODELS WITH RANDOM EFFECTS VIA THE GIBBS SAMPLER

3.1 Introduction

This chapter describes the Gibbs sampling algorithm that Karim (1991) and Zeger and Karim (1991) developed for Bayesian estimation in random effects generalized linear models. It first explains the logic of the algorithm. It then describes how to implement the technique to obtain Bayesian inferences on γ, T, and $\boldsymbol{u}$, which involves the use of an indirect sampling method called rejection sampling (Ripley, 1987) to generate variates from non-standard distributions. In addition, how the convergence of the Gibbs chains is assessed using Geweke (1992) and Raftery and Lewis' (1992) approaches is illustrated. The final section reports the results of a simulation study, which documents the performance of the SAS\IML code written to implement the Gibbs sampler and how well the algorithm works in analyses where the number of clusters, J, is less than the number of observations per cluster, n_j. The results are compared to those of Zeger and Karim, who established the validity of the algorithm in applications where J is large relative to the n_j's.

3.2 Logic of the Gibbs Sampler

Built on the work of Metropolis, Rosenbluth, Rosenbluth, and Teller (1953) and Hastings (1970), Gemen and Gemen (1984) employed Gibbs sampling in image

restoration. Later Gelfand and Smith (1990) and Gelfand et al. (1990) introduced this technique to the statistical community as a tool for fitting statistical models. Casella and George (1992) defined Gibbs sampling as "a technique for generating random variables from a ... distribution indirectly, without having to calculate the density" (p. 167). Its success lies in its ability to reduce "the problem of dealing simultaneously with a large number of intricately related unknown parameters ... into a much simpler problem of dealing with one unknown quantity at a time, sampling each from its *full conditional* distribution" (Gelman et al., 1995 p. 39). A full conditional distribution of a parameter is defined as its distribution conditional on the data and on the current values of all the other parameters in the model (Gilks, Best, & Tan, 1995). To obtain Bayesian inference in random effects generalized linear models, the technique reduces the complex tasks of computing the joint posterior distribution $p(\gamma,T,u|y)$ and the marginal posteriors $p(\gamma|y)$, $p(T|y)$ and $p(u|y)$ to a relatively straightforward task of sampling from the three full conditionals: $p(\gamma|T,u,y)$, $p(T|\gamma,u,y)$ and $p(u|\gamma,T,y)$. The strategy as well as the elegance of the technique can be revealed by re-expressing the joint posterior distribution as consisting of two components, a full conditional and a joint marginal distribution, in three different ways:

$$p(\gamma,T,u|y) = p(\gamma|T,u,y)p(T,u|y)\ , \qquad \textbf{(3.1)}$$

$$p(\gamma,T,u|y) = p(T|\gamma,u,y)p(\gamma,u|y), \qquad \textbf{(3.2)}$$

and

$$p(\gamma,\mathrm{T},\boldsymbol{u}\,|\,y) = p(\boldsymbol{u}\,|\,\gamma,\mathrm{T},y)p(\gamma,\mathrm{T}\,|\,y)\ . \qquad \textbf{(3.3)}$$

It is difficult to compute the joint marginal distributions above. For instance, the posterior distribution $p(\gamma,\mathrm{T}|\,y)$ is given by

$$p(\gamma,\mathrm{T}\,|\,y) = \frac{\prod_{j=1}^{J}\int\prod_{i=1}^{n_j} p(y_{ij}|\,\gamma,\boldsymbol{u}_j)p(\boldsymbol{u}_j|\,\mathrm{T})p(\gamma,\mathrm{T})\partial\boldsymbol{u}_j}{\int\prod_{j=1}^{J}\int\prod_{i=1}^{n_j} p(y_{ij}|\,\gamma,\boldsymbol{u}_j)p(\boldsymbol{u}_j|\,\mathrm{T})p(\gamma,\mathrm{T})\partial\boldsymbol{u}_j\partial\gamma\,\partial\mathrm{T}} \cdot \textbf{(3.4)}$$

However, as will be seen in the next sections, it is relatively easy to simulate from each of the three full conditionals above. The Gibbs sampler exploits this advantage. It employs the full conditional of each parameter to construct and drive a Markov chain which has, upon convergence, the joint distribution as its stationary or invariant distribution (Besag, Green, Higdon, & Mengersen, 1995). A property of Markov chain exploited here is that the probability of an event is conditionally dependent on a previous state. Let $k = 0, 1,...,K$ denote the k^{th} iteration in the Markov Chain simulation scheme. The iterative simulation scheme begins with some starting values $\gamma^{(0)}$, $\mathrm{T}^{(0)}$, and $\boldsymbol{u}^{(0)}$. Given $\mathrm{T}^{(0)}$ and $\boldsymbol{u}^{(0)}$ draw $\gamma^{(1)}$ from $p(\gamma|\mathrm{T}^{(0)},\boldsymbol{u}^{(0)},y)$, next draw $\mathrm{T}^{(1)}$ from $p(\mathrm{T}\,|\,\gamma^{(1)},\boldsymbol{u}^{0)},y)$, and then conclude an iteration cycle by drawing $\boldsymbol{u}^{(1)}$ from $p(\boldsymbol{u}|\gamma^{(1)},\mathrm{T}^{(1)},y)$. Gemen and Gemen showed that after the sampler has been run for enough iterations to achieve convergence, $\gamma^{(k)}$, $\mathrm{T}^{(k)}$, and $\boldsymbol{u}^{(k)}$ can be regarded as random variates from $p(\gamma,\mathrm{T},\boldsymbol{u}|y)$, regardless of the starting values chosen.

The joint distribution or any lower dimensional marginal distribution can be

approximated by the empirical distribution of r subsequent successive draws obtained from the Gibbs chain after convergence, where $r = 1$ to R is chosen to provide sufficient precision to the empirical distributions of interest (Karim, 1991). If a conditional distribution has a closed form, such as T in random effects generalized linear models, the empirical distribution can also be estimated by

$$p(\mathrm{T}\,|\,y) = \frac{1}{R}\sum_{r=1}^{R} p(\mathrm{T}\,|\,\gamma^{(r)},\boldsymbol{u}_j^{(r)}) \ . \qquad \textbf{(3.6)}$$

The next section describes how the Gibbs sampler is applied to obtaining Bayesian inferences on the various parameters of interest.

3.3 Implementation of the Gibbs Sampler

The implementation of the Gibbs sampler requires first the derivation of each of the conditional distributions from the joint posterior distribution $p(\gamma,\mathrm{T},\boldsymbol{u}|y)$:

$$\begin{aligned} p(\gamma,\mathrm{T},\boldsymbol{u}\,|\,y) \propto & \prod_{j=1}^{J}\prod_{i=1}^{n_j} \frac{1}{1+\exp^{-\eta_{ij}}}^{y_{ij}} \frac{\exp^{-\eta_{ij}}}{1+\exp^{-\eta_{ij}}}^{1-y_{ij}} \cdot \\ & |\ \mathrm{T}\ |^{-1/2}\exp[-\frac{1}{2}\boldsymbol{u}_j^{\mathrm{T}}\mathrm{T}^{-1}\boldsymbol{u}_j] \cdot p(\gamma,\mathrm{T}) \ . \end{aligned} \qquad \textbf{(3.7)}$$

The strategy is to pick out the term(s) in the joint density **(3.7)** which involve the relevant parameter (Gilks, 1996).

3.3.1 Full Conditional Distribution of γ--$p(\gamma \mid T, \boldsymbol{u}, y)$

Assuming the priors $p(\gamma) \propto$ *constant* and $p(T) \propto$ *constant*, and that the two priors are independent, i.e.,

$$p(\gamma, T) = p(\gamma)p(T) , \qquad \textbf{(3.8)}$$

$p(\gamma \mid T, \boldsymbol{u}, y)$ is independent of T. The conditional distribution can thus be stated as $p(\gamma \mid \boldsymbol{u}, y)$. Given the values $\boldsymbol{u}^{(k)}$ for $\boldsymbol{u}$ at iteration k, with reference to the joint density **(3.7)**, $p(\gamma \mid \boldsymbol{u}^{(k)}, y)$ is proportional to the likelihood function

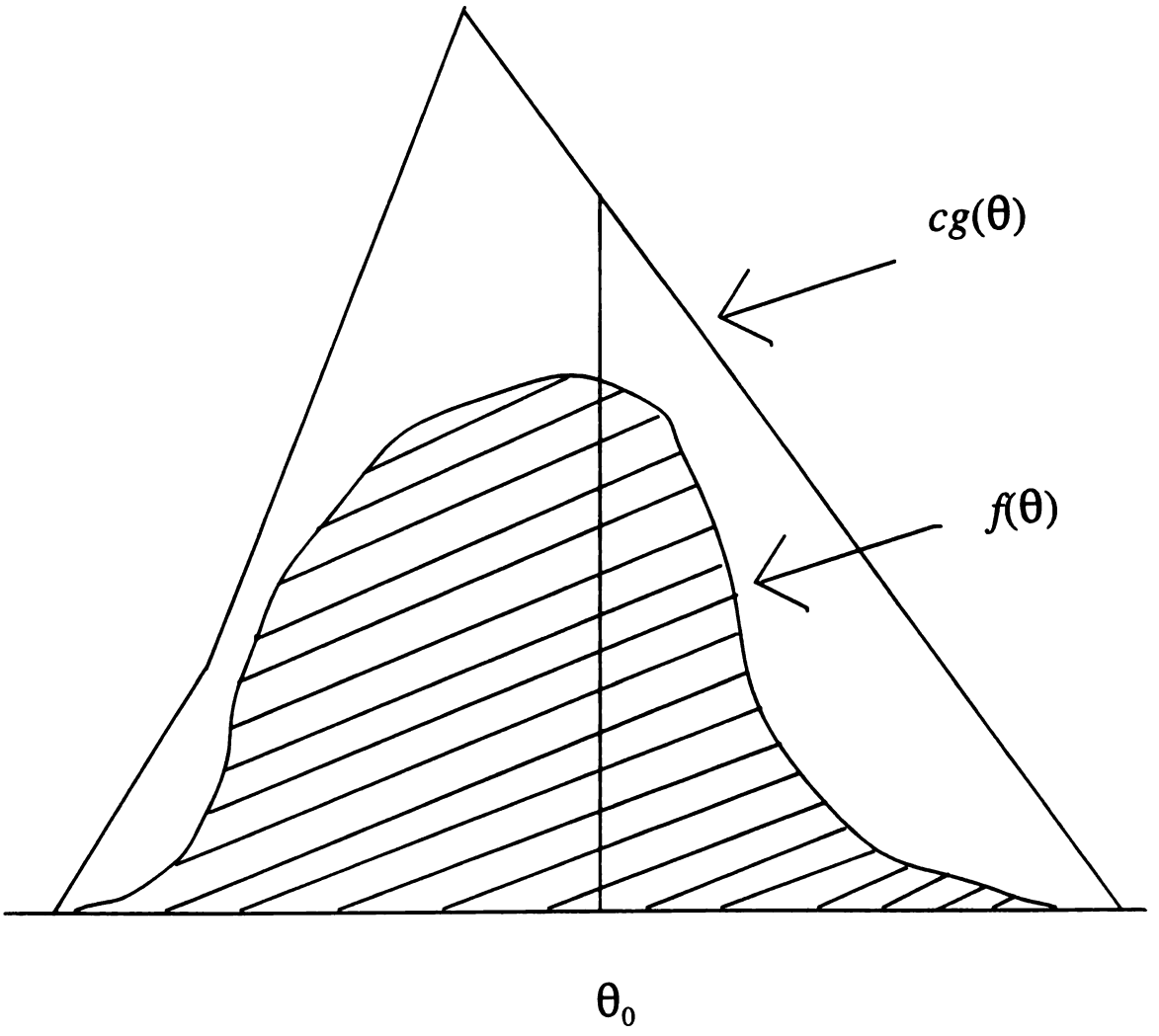
$$p(\gamma \mid \boldsymbol{u}^{(k)}, y) \propto \prod_{j=1}^{J} \prod_{i=1}^{n_j} \left(\frac{1}{1 + \exp^{-(x_{ij}^T \gamma + z_{ij}^T u_j^{(k)})}} \right)^{y_{ij}} \cdot \left(\frac{\exp^{-(x_{ij}^T \gamma + z_{ij}^T u_j^{(k)})}}{1 + \exp^{-(x_{ij}^T \gamma + z_{ij}^T u_j^{(k)})}} \right)^{1 - y_{ij}} \cdot \qquad \textbf{(3.9)}$$

As no standard algorithm exists for generating random variates from the non-standard conditional distribution **(3.9)**, Zeger and Karim (1991) adopt the rejection\acceptance sampling algorithm (Ripley, 1987). To illustrate the logic of the technique, the procedure for generating variates having a non-standard probability density function (p.d.f.) $f(\theta)$ and a cumulative distribution function (c.d.f.) $F(\theta)$ is outlined here.

To begin with, a different p.d.f $g(\theta)$ known as an envelope function with c.d.f. $G(\theta)$, which resembles $f(\theta)$ and is easy to sample from, is chosen. Then random variates are generated from $g(\theta)$. If the random variates drawn pass an acceptance\rejection test, which will soon be illustrated, the variates will be accepted as belonging to $f(\theta)$. The algorithm requires that the envelope function, $g(\theta)$, dominate $f(\theta)$. To ensure its coverage

over $f(\theta)$, $g(\theta)$ is always multiplied by a constant c. The more closely $g(\theta)$ imitates $f(\theta)$, the lower the rejection rate will be. Figure 3.1, taken from Gelman et al. (1995, p.304), shows how $cg(\theta)$ envelopes $f(\theta)$. The vertical line indicates that a random variate $\theta_{.}$ is sampled from $g(\theta)$, with a realization θ_0.

Figure 3.1 Illustration of Rejection Sampling (Taken from Gelman et al., 1995, p. 304)



The ratio, $r(\theta)$, of the height of the lower curve $f(\theta_0)$ to the height of the upper curve $cg(\theta_0)$ gives the probability that θ_0 will behave according to $f(\theta)$. To test if θ_0 is to be accepted, a number U is drawn at random from the unit interval [0,1]. If $U \le r(\theta)$, θ_0 is accepted. If $U > r(\theta)$, θ_0 is rejected. If the random variate is rejected, another draw from $g(\theta)$ and the unit distribution will be made, followed by another acceptance/rejection test. The validity of the algorithm lies in the equivalence of $G(\theta_0)$ conditional on $U \le r(\theta)$ to $F(\theta_0)$, as shown by Dagpunar (1988).

The envelope function chosen by Zeger and Karim to cover the full conditional distribution **(3.9)** is the Gaussian density $N(\hat{\gamma}^{(k)}, 2 \cdot V_{\gamma}^{(k)})$, where $\hat{\gamma}^{(k)}$ maximizes $p(\gamma \mid \boldsymbol{u}^{(k)}, y)$ and $V_{\gamma}^{(k)}$ is the inverse Fisher information at the k^{th} iteration of the Gibbs sampler. Given the values of $\boldsymbol{u}^{(k)}$, the task of estimating $\hat{\gamma}^{(k)}$ and $V_{\gamma}^{(k)}$ is simplified because $z_{ij}^{\mathrm{T}} \boldsymbol{u}_j^{(k)}$ has become a vector of offsets, or explanatory variables with known coefficients (Gelman et al., 1995). Model **(2.7)** therefore reduces to a fixed effects generalized linear model, which can be estimated via iteratively weighted least squares with $y_{ij}^{(k)*}$ being the linearized dependent variable and w_{ij} being the weight. How they can computed is given in the equations **(2.23)** and **(2.24)**. Let $\boldsymbol{y}_j^{*(k)} = (y_{1j}^{*(k)}, \ldots, y_{n_j j}^{*(k)})$, $\boldsymbol{X}_j = (\boldsymbol{x}_{1j}^{\mathrm{T}}, \ldots, \boldsymbol{x}_{n_j j}^{\mathrm{T}})$, $\boldsymbol{Z}_j = (\boldsymbol{z}_{1j}^{\mathrm{T}}, \ldots, \boldsymbol{z}_{n_j j}^{\mathrm{T}})$, $\boldsymbol{W}_j^{(k)} = \mathrm{diag}\{p_{ij}^{(k)}(1-p_{ij}^{(k)})\}$, then at the k^{th} iteration of the Gibbs chain,

$$\begin{aligned} \hat{\gamma}^{(k)} &= \Big(\sum_{j=1}^{J} \boldsymbol{X}_j^{\mathrm{T}} \boldsymbol{W}_j^{(k)} \boldsymbol{X}_j\Big)^{-1} \Big(\sum_{j=1}^{J} \boldsymbol{X}_j^{\mathrm{T}} \boldsymbol{W}_j^{(k)} \boldsymbol{y}_j^{(k)^*}\Big), \\ V_{\gamma}^{(k)} &= \Big(\sum_{j=1}^{J} \boldsymbol{X}_j^{\mathrm{T}} \boldsymbol{W}_j^{(k)} \boldsymbol{X}_j\Big)^{-1}. \end{aligned} \qquad \textbf{(3.10)}$$

To ensure that the envelope function covers the full conditional **(3.9)**, Zeger and Karim (1991) inflate the variance $V_{\hat{\gamma}}^{(k)}$ by a factor of 2 and multiply the envelope function with a constant c, which matches the mode of the envelope function and the full conditional **(3.9)**. Let the full conditional **(3.9)** be denoted by $f(\gamma)$ and the envelope function $N(\hat{\gamma}^{(k)}, 2 \cdot V_{\hat{\gamma}}^{(k)})$ by $g(\gamma)$, c is then computed as the ratio $f(\hat{\gamma}^{(k)})/g(\hat{\gamma}^{(k)})$. Using a Cholesky factorization of $V_{\hat{\gamma}}^{(k)}$, a random variate γ^* can be generated. Then a rejection\acceptance test involves evaluating the full conditional **(3.9)** and the envelope function at γ^* by computing the ratio $f(\gamma^*)/(c \cdot g(\gamma^*))$ is implemented. The decision rule is that if the ratio is greater than or equal to a random variate U generated from the uniform distribution [0,1], γ^* is accepted as a random variate belonging to the full conditional **(3.9)** with a probability equal to the ratio and $\gamma^{(k+1)} = \gamma^*$. Otherwise, draw another variate from the envelope function and subject it to the rejection\acceptance test again.

3.3.2 Full Conditional Distribution of T--$p(\mathrm{T} \mid \gamma, u, y)$

Assuming u_j's are independent normal with mean 0 and variance T and a uniform prior for T, i.e., $p(\mathrm{T}) \propto$ constant, the full conditional distribution is independent of γ and y. It thus can be expressed as $p(\mathrm{T} \mid u^{(k)})$, which is proportional to

$$p(\mathrm{T} \mid u^{(k)}) \propto \prod_{j=1}^{J} |\ \mathrm{T}\ |^{-1/2} \exp[-\frac{1}{2} u_j^{(k)\mathrm{T}} \mathrm{T}^{-1} u_j^{(k)}] \ . \qquad \textbf{(3.11)}$$

Another choice of prior for $p(\mathrm{T})$, which Zeger and Karim use, is the Jeffrey's prior $|\mathrm{T}|^{-(q+1)/2}$. More recent work on Bayesian analysis for hierarchical models (e.g., Seltzer, Wong, & Bryk, 1996), however, has found that the choice may cause a problem. The

problem can be illustrated with the case when the variance term, τ, is a scalar, that is q = 1. The corresponding Jeffrey's prior is $p(\tau) \propto 1/\tau$, $\tau > 0$. When τ is close to zero, $1/\tau$ becomes infinitely large and "a spike at $\tau = 0$" (Seltzer et al., 1996, p. 139) in the posterior distribution for the parameter may occur. It will result in a standstill in the updating scheme of the Gibbs sampler. I encountered this problem in the simulation runs and the problem went away when I adopted the uniform prior, i.e., $p(\tau) \propto$ *constant*.

To obtain random draws from **(3.11)**, it is convenient to work with the conditional posterior distribution of $\mathbf{T}^{-1}$, given $\boldsymbol{u}$, which is a Wishart distribution with parameters

$$\mathbf{S}^{(k)} = \sum_{j=1}^{J} \boldsymbol{u}_j^{(k)} \boldsymbol{u}_j^{(k)\mathrm{T}} , \qquad \textbf{(3.12)}$$

and J-q-1 degrees of freedom, where q is equal to the dimension of T. As one can use a standard algorithm for sampling from the full conditional of $\mathbf{T}^{-1}$ here, rejection sampling is not needed. $\mathbf{T}^{(k+1)}$ can then be obtained by inverting the values of $\mathbf{T}^{-1}$ sampled. Specifically, $\mathbf{T}^{-1}$ is equal to $\mathbf{H}^{(k)\mathrm{T}}\mathbf{WIH}^{(k)}$, where $\mathbf{S}^{(k)-1} = \mathbf{H}^{(k)\mathrm{T}}\mathbf{H}^{(k)}$ and **WI** is a standardardized Wishart variate with J-q-1 degrees of freedom (Odell and Feiveson, 1966).

3.3.3 Full Conditional Distribution of *u*--$p(\boldsymbol{u} \mid \gamma, \mathrm{T}, y)$

Given the values $\gamma^{(k)}$ and $\mathbf{T}^{(k)}$, the full conditional distribution $p(\boldsymbol{u} \mid \gamma^{(k)}, \mathbf{T}^{(k)}, y)$ is proportional to

$$p(\boldsymbol{u} \mid \gamma^{(k)}, T^{(k)}, y) \propto \prod_{j=1}^{J} \prod_{i=1}^{n_j} \left(\frac{1}{1 + \exp^{-(x_{ij}^T \gamma^{(k)} + z_{ij}^T u_j)}} \right)^{y_{ij}} \cdot \left(\frac{\exp^{-(x_{ij}^T \gamma^{(k)} + z_{ij}^T u_j)}}{1 + \exp^{-(x_{ij}^T \gamma^{(k)} + z_{ij}^T u_j)}} \right)^{1 - y_{ij}} \cdot \mid T^{(k)} \mid^{-1/2} \exp[-\frac{1}{2} u_j T^{(k)^{-1}} u_j] \ . \quad \textbf{(3.13)}$$

(3.13) is a non-standard distribution, so Zeger and Karim (1991) again use the rejection sampling algorithm. For each cluster j, the envelope function employed is $N(\hat{u}_j^{(k)}, 2 \cdot V_{\hat{u}j}^{(k)})$. Using iterative weighted least squares, at the k^{th} iteration of the Gibbs sampler, the mode $\hat{u}_j^{(k)}$ for cluster j

$$\hat{u}_j^{(k)} = (Z_j^T W_j^{(k)} Z_j + T^{-1^{(k)}})^{-1} Z_j^T W_j^{(k)} (y_j^{*(k)} - X_j \gamma^{(k)}) \ , \quad \textbf{(3.14)}$$

and its curvature

$$V_{\hat{u}_j}^{(k)} = (Z_j^T W_j^{(k)} Z_j + T^{-1^{(k)}})^{-1} \ , \quad \textbf{(3.15)}$$

can be estimated. To ensure coverage, the variance of the envelope function is inflated by a factor of 2, and the function itself is multiplied by a constant c_j, which matches the modes of the full conditional and the envelope function for cluster j. For each cluster, a random variate u_j^* is drawn from the envelope function. A rejection\acceptance test that computes and compares the ratio $f(u_j^*)/(c_j \cdot g(u_j^*))$, where $f(u_j)$ is the full conditional for cluster j and $g(u_j)$ is the envelope function, to a random variate U generated from the uniform distribution [0,1]. If $U \geq f(u_j^*)/(c_j \cdot g(u_j^*))$, $u_j^{(k+1)} = u_j^*$, otherwise repeat the sampling procedure.

3.3.4 Starting Values and Convergence Diagnostics

The starting values of the Gibbs sampler for $\boldsymbol{u}_j$'s and $\mathbf{T}$ are set to be zeros and the identity matrix respectively by Karim (1991). No starting values for γ are needed as the Gibbs sampler starts with sampling from the full conditional distribution of γ. To assess convergence at iteration k, Zeger and Karim (1991) employ Q-Q plots in which the last specified number of draws, say m, are plotted against the preceding m draws. In the present investigation, the convergence diagnostics developed by Geweke (1992) and Raftery and Lewis (1992) and automated in the software Convergence Diagnosis and Output Analysis Software for Gibbs Sampling Output (CODA) (Best et al., 1995; Cowles, 1994) are used to monitor the convergence of the Gibbs chains. Besides the availability of the computer code, the two diagnostics are chosen because 1) the methods are theoretically motivated, 2) each of them requires one sampler run, and 3) the conclusion drawn from each diagnostic could be used to compare with one another to check for the reliability of the diagnostic results.

To briefly explain and illustrate the two approaches, I apply the diagnostics to a chain of 800 simulated values of τ generated to monitor the convergence of the chain. The parameter value for τ is 0.250.

The first diagnostic used is Geweke's (1992) convergence diagnostic, which is based on a standard time-series method. For each variable, the chain is divided into two "windows" containing different fractions of the chain, for example, the first 10% and the last 50%. If the whole chain is stationary, the means of the values early and late in the sequence should be similar. Geweke's approach involves computing a convergence

diagnostic Z, which is the difference between the 2 means in the two different windows divided by the asymptotic standard error of their difference. As the length of the chain $K \to \infty$, the sampling distribution of $Z \to N(0,1)$ if the chain has converged. Thus values of Z which fall in the extreme tails of a standard normal distribution may imply that the chain has not fully converged. Table 3.1 gives the CODA output for Geweke's convergence diagnostic.

Table 3.1: CODA Output for Geweke Convergence Diagnostic

GEWEKE CONVERGENCE DIAGNOSTIC (Z-score):
==========

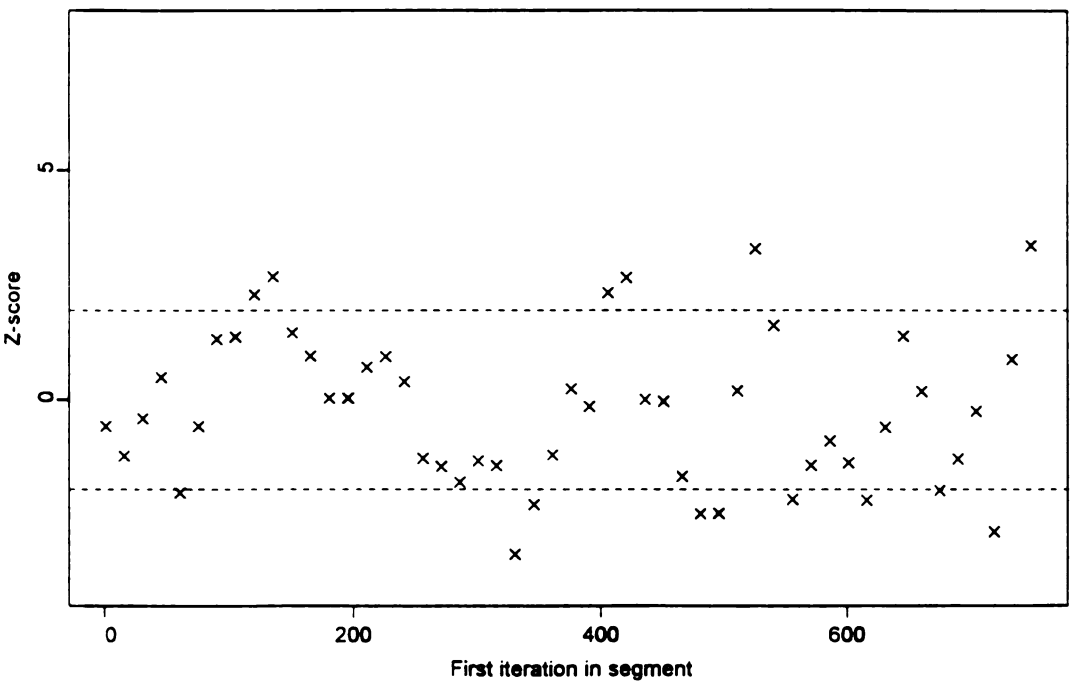
Iterations used = 1:800
Thinning interval = 1
Sample size per chain = 800

Fraction in 1st window = 0.1
Fraction in 2nd window = 0.5

Variable	Convergence Diagnostic Z
τ	-0.594

As Table 3.1 shows, the number of iterates used is 800. The diagnostic uses every one of the 800 iterates, thus the thinning interval is equal to 1. Sometimes instead of using every iteration, one may "thin" the chain by choosing to save and use every t^{th} iteration, where $t > 1$. The option is useful in runs when there is high correlation between consecutive iterations (Raftery & Lewis, 1996). Here I follow Zeger and Karim (1991) in using every iteration and a single chain. The two "windows" contain the first 10% and the last 50% of the iterates respectively. Given that the Z-score is only -0.594, no evidence against convergence is established. According to Geweke (1992), a score in excess of 4 indicates problems. One can also use plots to illustrate the diagnostic. In essence, the approach computes and plots different Z-scores for different segments of the chain. A large portion of Z-scores falling outside the 95% confidence interval for a $N(0,1)$ distribution will suggest possible convergence failure. Figure 3.2 gives the plot of the analysis and the 95% confidence interval. As shown in the figure, most of the scores are within or cluster near the confidence interval (the two broken lines), therefore, no convergence failure is suggested.

Figure 3.2 Geweke's Convergence Diagnostic



Raftery and Lewis' (1992) approach, based on two-state Markov chain theory, as well as standard sample size formulas involving binomial variance, is an alternative convergence diagnostic. The method detects convergence as well as provides a way of bounding the variance of the estimates of quantiles of functions of parameters. Table 3.2 gives a slightly edited CODA output of the analysis of the same chain of 800 simulated values used earlier. Several letter symbols of the original output were changed to avoid confusions with the ones already employed in the thesis.

Table 3.2: CODA Output for Raftery and Lewis Convergence Diagnostic

RAFTERY AND LEWIS CONVERGENCE DIAGNOSTIC:

==========

Sample size per chain = 800

Quantile = 0.025
Accuracy = +/- 0.02
Probability = 0.9

Chain: c:/coda/diagnose
==========

VARIABLE	Thin (t)	Burn-in (B)	Total (N)	Lower bound (Nmin)	Dependence factor (D)
τ	1	5	253	165	1.53

In the output, the sample size per chain indicates the number of iterates used, which is again 800. This diagnostic specifies the convergence criteria to be estimating the 0.025 quantile of the posterior distribution of τ with a precision of ± 0.02 with probability 0.9. The burn-in value indicates how many initial iterations can be discarded. The small burn-in value given in Table 3.2, B = 5, suggests the chain has converged almost immediately to τ. In order to estimate the 2.5th percentile of the posterior distribution to the specified accuracy and probability, i.e., ± 0.02 and 0.9 respectively, the result recommends that an estimated 254 iterations (with a minimum of 165) out of the 800 iterations are needed. The last column reports the dependence factor, which is the ratio of N to Nmin. Values of D much greater than 1.0 indicate high within-chain correlations and probable convergence failure (Raftery and Lewis (1996) suggest that D > 5.0 often indicates problems). The result here (D = 1.53) indicates convergence is achieved.

3.4 A Simulation Study and Results

A small simulation study was done to assess the accuracy of the SAS/IML code and how well it would suit the present analysis. The structure of the data sets generated basically followed that of the simulations carried out by Rodriguez and Goldman (1993) and Yang (1995), in which τ was a scalar. The model specifications of the simulated data were similar to the over-simplified model **(2.6)** discussed earlier. The model had one predictor at each level, denoted by $lvl1pred_{ij}$ and $lvl2pred_j$ for the level-1 and level-2 predictors. However, here only the level-1 intercept but not the regression coefficient was

modeled using $lvl2pred_j$. The model could be expressed as:

$$\eta_{ij} = (1 \;\; lvl2pred_j \;\; lvl1pred_{ij}) \begin{bmatrix} \gamma_{00} \\ \gamma_{01} \\ \gamma_{10} \end{bmatrix} + [1][u_{0j}] . \qquad \textbf{(3.16)}$$

Two kinds of data sets, each had $J = 40$ and $n_j = 100$, were generated. The number of clusters and observations per cluster was chosen to be similar to those of the TSAP in mathematics, which had 42 states and 100 schools per state on average. In both data sets, $\gamma_{00} = -.796$, $\gamma_{01} = \gamma_{10} = 1$. They differed in their value of τ. One was equal to 0.250 and the other was equal to 1. The two different values were chosen to examine how well the algorithm works with moderate as well as large between-cluster variance. As the algorithm is very computationally intensive, the number of replications was chosen to be 50, half of what Zeger and Karim (1991) ran. It is important to note that the goals of this simulation, as stated in the beginning of this section, was different from Zeger and Karim's, which was to establish the statistical properties of the Gibbs sampler algorithm.

The Gibbs sampler was run for 2,800 to 5,000 iterations until convergence. Table 3.3 lists the values of the true parameter θ, and reports the mean and standard deviation of $\hat{\theta}$, the square root of the average variance of the estimates, and the coverage of the nominal 90% Bayes interval from the 5th to 95th percentile based on 200 additional simulated values generated after convergence. 200 iterates were used as it would facilitate the comparison of the results of this simulation with those of Zeger and Karim (1991),

who used the same number of iterates to compute the posterior characteristics of the various parameters.

Table 3.3: Results of Simulation Study

θ =	γ_{00}	γ_{01}	γ_{10}	τ
Case 1				
θ (true value)	-0.796	1.000	1.000	0.250
mean($\theta \mid y$)	-0.763	1.042	0.991	0.287 (mode = 0.250)
std($\theta \mid y$)	0.187	0.233	0.152	0.104
$(mean(var((\theta \mid y))))^{1/2}$	0.163	0.201	0.154	0.096
Coverage of nominal 90% interval	84	82	94	88
Case 2				
θ (true value)	-0.796	1.000	1.000	1.000
mean($\theta \mid y$)	-0.778	1.042	1.016	1.211 (mode = 1.070)
std($\theta \mid y$)	0.302	0.372	0.144	0.359
$(mean(var((\theta \mid y))))^{1/2}$	0.270	0.334	0.159	0.349
Coverage of nominal 90% interval	86	84	96	82

The results are similar to those of Zeger and Karim (1991). In their simulation, they reported a slight bias in the intercept estimate while the other fixed effects were approximately unbiased. Here all fixed effects are approximately unbiased. Zeger and

Karim reported that the posterior means for the random effects variances were positively biased by 20%-30%. They attributed the bias to the long tail of the marginal distribution, which was alleviated when posterior modes were used instead of means to indicate central tendency. Here the positive bias ranges from 15% to 22% and the bias is alleviated by using posterior modes. The coverage probabilities of the nominal 90% Bayes posterior intervals ranged from 80% to 100% in Zeger and Karim's simulation study and they range from 82% to 96% here. Given the number of trials was 50, the standard deviation of the proportion of coverage is *sqrt*(.9(1-.9)/50) = .04. Thus two standard deviations below and above the 90% coverage ranged from .82 to .98. The results obtained here, 82% to 96%, gave another indication that the code and the algorithm were performing well.

Both simulations show the algorithm yields reasonable inferences, which suggests that the algorithm works well in analyses where the number of clusters, J, is less than the number of observations per cluster, n_j as well and can be applied to analyze the TSAP in mathematics data to study the distribution of learning opportunities. The next chapter reports a use of the tested code to perform Bayesian analysis of models studying access to opportunities to learn.

CHAPTER 4

BAYESIAN HIERARCHICAL ANALYSES OF ACCESS TO EIGHTH-GRADE ALGEBRA

4.1 Introduction

In this chapter, the estimation approach and algorithm from Chapter 3 are implemented to analyze the data collected under the 1992 TSAP in mathematics and state background characteristics data made available by CCSSO. The purpose of the analysis is to study the distribution of learning opportunities in advanced mathematics in Grade 8 among schools across 42 U.S. states. I first provide a description of the data sets and the procedures used. Then the selection and the construction of relevant variables and measures are presented in conjunction with the various research hypotheses. The final sections describe and discuss the findings and their implications.

4.2 Data, Procedures, and Measures

4.2.1 Data Description

In TSAP, 44 states volunteered to participate. Within each state, on average, 100 public schools were selected at random. Within each school, approximately 30 eighth graders were then sampled. More specifically, eligible schools were first stratified and selected according to grade 8 enrollment, urbanicity, percentage of African and Hispanic American students, and median household income (for a description of the sample design and selection, see Mohadjer, Rust, Smith and Severynse, 1993). A systematic sample of

students was then drawn from each selected school. In the survey, the students were administered a test to assess their mathematics proficiency. Personal and school background information was solicited from the students, their teachers, and their school principals or administrators.

The present investigation employed primarily the school response data on school characteristics and policies collected from the principals or other administrators. After listwise deletion of cases with missing values, complete data were available for 3,525 schools in 42 states. About 5% of schools in the total sample (3,725) and two territories were removed as a result. The two territories, each with six schools, were excluded as they had no data on the availability of algebra. A comparison of the descriptive statistics of the total sample and the actual sample used in the analyses revealed little difference. Another data set used in conjunction with TSAP was data on state profile statistics compiled by CCSSO, for which there were no missing data (CCSSO, 1993). In the analysis, the two data files were merged.

4.2.2 Procedures

Using the TSAP and state background data, I constructed measures of the central concepts of this study for the selected schools and participating states in the sample. Reliability analyses and the less computationally intensive penalized quasi-likelihood estimation approach implemented via the EM algorithm (Raudenbush, 1993) were used in scale construction.

With the log odds of the offering of high school algebra as the outcome, I first

formulated an unconditional model that included 1) a within-states model in which the outcome for a given school in a given state was seen as varying around the mean of the outcome for that state and 2) a between-states model in which the mean for the state on that outcome was seen as varying randomly around the grand mean for the outcome. The model estimated the unconditional variance in the outcome that lies between states.

To test hypotheses, I then expanded the model to include school- and state-level predictors, which consisted of school composition, setting, eighth-grade enrollment, state poverty level and educational expenditure. In the analysis, all coefficients that represented the effects of school-level predictor variables were assumed not to randomly vary among the 42 states. Systematic state-to-state variation of the racial and social composition effects on the offering of algebra associated with different levels of public educational investment was postulated. The model specification regarding between-state variation of the effects of the various school-level predictors was guided by the principle of parsimony and based on the reasoning that the effects on the outcome due to school size and urbanicity are similar across the 42 states. And after controlling for the state educational investment and poverty level, it is likely that there will be no significant unique ethnicity and social composition effects associated with individual states. An exploratory run using HLM2/3 (Bryk et al., 1996) found non-significant random variation in the ethnicity and social composition effects, controlling for the two state-level predictors.

In addition, all predictors in the within- and between-state model were centered around their means. As a result, the mean logits of offering algebra for individual states were adjusted for the various predictors.

4.2.3 Measures

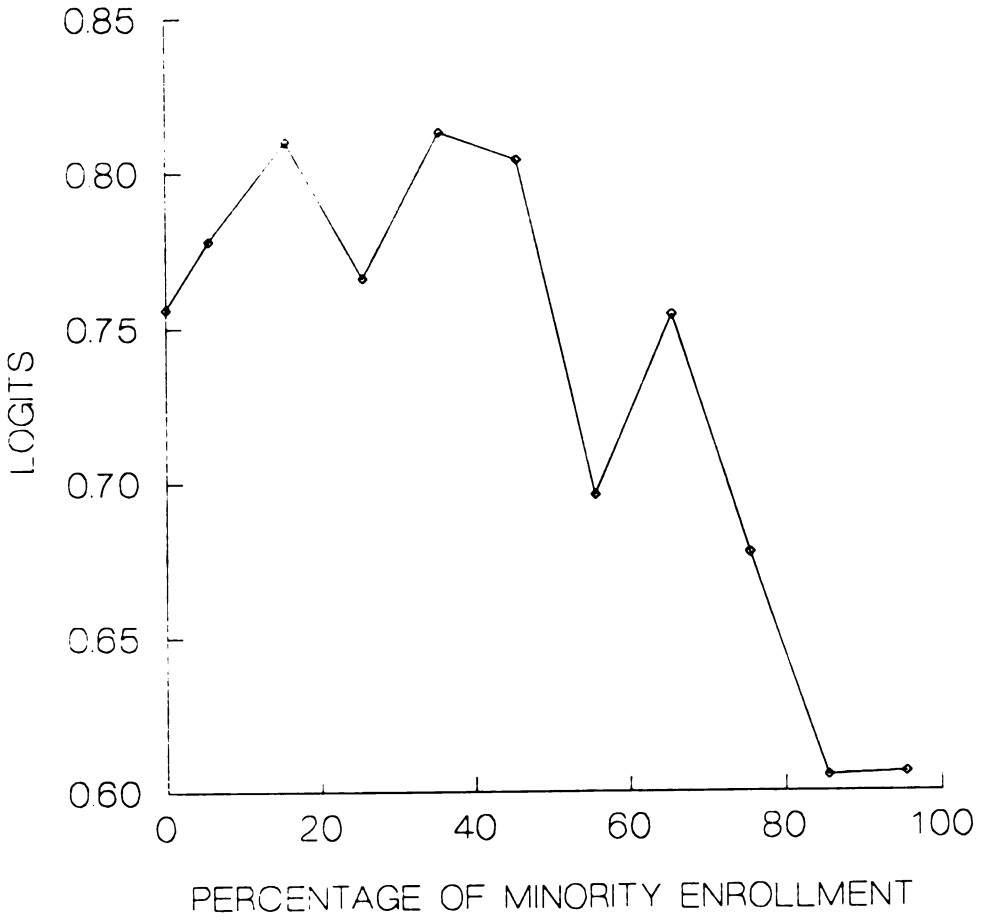
A central focus of this research is to describe and study the availability of high school algebra in the 42 states and the extent to which the availability depends on various demographic composition, setting, structural, and state economic and educational financial factors. To measure the availability of the course, an indicator variable was used. Schools received a score of 1 if the course was offered and a 0 otherwise. The mean for the variable was .76, indicating that 76% of the 3,525 schools offered the course (the standard deviation was .43). 846 schools in the sample did not include algebra in their eighth-grade mathematics curriculums.

Two measures of the demographic composition were the minority enrollment and the SES of a school. Schools in which African and Hispanic American students made up more than 50% of their enrollments were represented by an indicator variable and coded as high minority enrollment schools. The variable was constructed from the continuous variable of percentage of minority students in a school. The decision to adopt the dummy coding scheme was based on an examination of various plots of predicted logits of the probability of offering algebra against the midpoints of grouped percentages of minority enrollment, as suggested by DeMaris (1992). In addition, an exploratory model was fit with an alternative coding scheme for the independent percent minority variable.

Eleven groups of percentages (0, 1-10, 11-20, 21-30, 31-40, 41-50, 51-60, 61-70, 71-80, 81-90, and 91-100) were formed by breaking the range of the percentages of minority students in schools. All but the no minority enrollment group, which was chosen to be the reference group, were used to model the log odds of offering algebra using

HLM2\3 (Bryk et al., 1996). Thus the model had ten dummy predictor variables. After the run, I plotted the estimated logits for all eleven categories of percentages against the eleven midpoints of the groups. The estimated logits for all but the reference group were computed by adding the regression coefficient of each group to the intercept, with the intercept being the estimated logit for the no minority group. Figure 4.1 gives the plot. The result shows that higher percentage of minority enrollment has a generally negative relationship to the logits of offering algebra after the 40th to 50th percentile category. In addition, I plotted the estimated logits against the logarithm and the logits of the midpoints of the various groups. It was found that the two transformations of the independent minority enrollment variable did not result in linearity in the logit of offering algebra. To explore an alternative coding scheme, I ran a model with percentage of minority enrollment and its quadratic term as predictors. The result of the run showed that the quadratic term did not achieve a conventional level of significance, and no quadratic relationship was suggested. Therefore the analysis used an indicator variable indicating schools with high minority enrollment, which allowed a straightforward interpretation as well.

Figure 4.1. Logits Plotted Against the Midpoints of the Categories of Percentages of Minority Enrollment

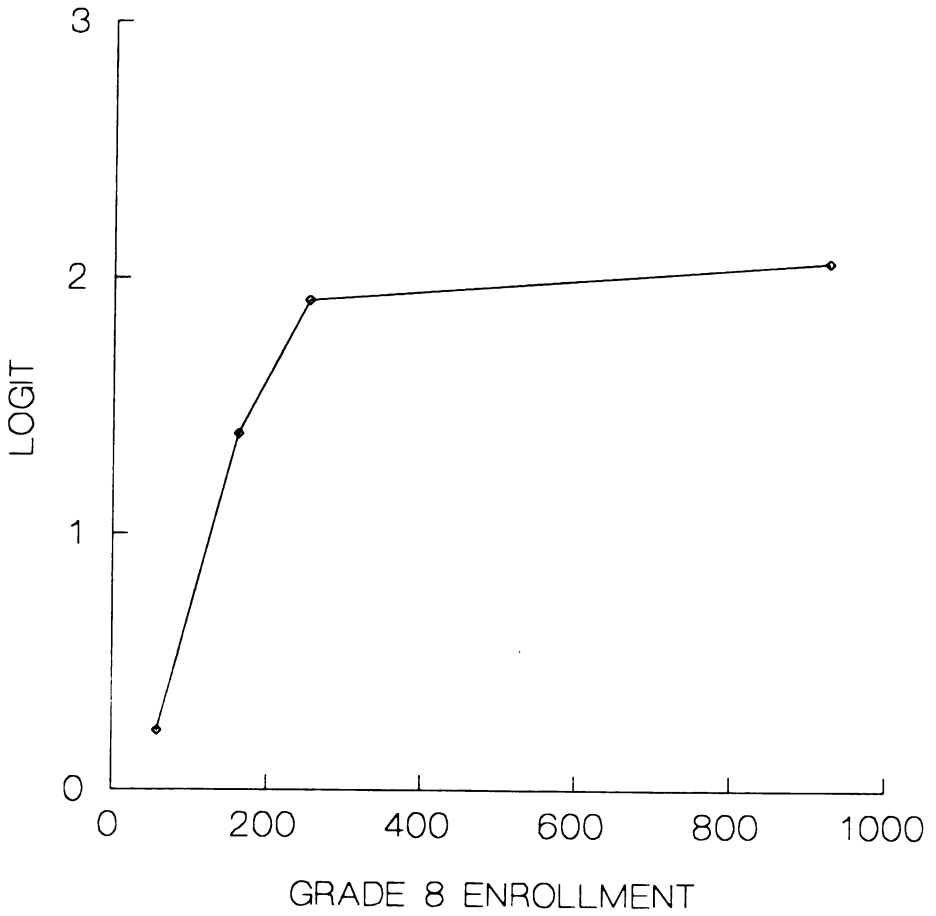


The measure of the SES of schools was a 3-item scale. Two items in the scale covered the home background of students of a school, and the third was the percentage of students who were in the free lunch program. The home background items were based on students' reports about the educational level of their parents and their access to reading materials (receiving a newspaper regularly, an encyclopedia in the home, more than 25 books in the home, and receiving regular magazines). To create the SES scale, I first averaged the responses of the students within each school to obtain the school means for the two measures on home background. Z scores were then created for the two aggregated home background measures and the percentages of students receiving free lunches. Finally, the scale was constructed as the average of the three standardized scores. The scale had a Cronbach's alpha of 0.76.

Grade 8 enrollment was measured by the actual number (in hundreds) of students enrolled. Past research suggests that the positive relationship between school size and program comprehensiveness is nonlinear (e.g., Monk & Haller, 1993). The relationship weakens as the size of the school increases. The finding warrants a check of the tenability of the assumption of linearity in the logit of the outcome. To proceed, I broke the range of the enrollment variable into groups, following the approach of DeMaris (1992) and Hosmer and Lemeshow (1987). The values of the variable ranged from 0.01 to 16.07. Grouping based on quartiles of the distribution of the variable yielded four categories with midpoints at 0.585, 1.615, 2.525, and 9.25. I then created three indicator variables to represent three of the four categories and entered them as predictors to model the log odds of offering algebra. To compute the estimated logits for each category, I added

successively the regression coefficient for each indicator variable to the intercept, which was the estimated logit for the reference group. Finally, I plotted them against the midpoints of the four categories. Figure 4.2 displays the plot of logits against the various midpoints. It shows obvious departure from a straight line and reveals nonlinearity in the relationship. As a result, a quadratic term was created to appropriately model the trend.

Figure 4.2: Plot of Logits Against Midpoints of Grouped Categories of Enrollment



The measure of the setting of the school was represented by three indicator variables showing whether or not a school was located in an urban, suburban, or rural setting. In the sample, 23% were urban schools, 53% were suburban schools, and 24% were rural schools.

At the state level, two measures employed were the percent of children in poverty and the state educational expenditure per pupil. CCSSO compiled the two measures for each state for 1990, which was relatively close in time to 1992, the year the data on algebra offering were gathered (CCSSO, 1993). Percent of children in poverty is defined as "the percentage of related children under age 18 who live in families with incomes below the U.S. poverty threshold" (The Annie E. Casey Foundation, 1994, p. 159). The state educational expenditure measures how much money (in hundreds) states allocate for each student, and indexes state efforts in funding public education. Percent of children in poverty is negatively correlated with state educational expenditure per pupil (r = -.41).

Table 4.1 shows the means, standard deviations, ranges, and variable names for each of the variables in this analysis.

Table 4.1: Descriptive Statistics

Variables	Range	Mean	SD
<u>School-level Data</u> (3,525 schools)			
Offering 8th Grade Algebra for High School Credits	0 = no 1 = yes	0.76	0.43
Schools with Minority (Hispanic and African Americans) in Excess of 50%	0 = no 1 = yes	0.14	0.35
School SES	(-3.42, 2.08)	0.00	0.82
Schools in Urban Setting	0 = no 1 = yes	0.23	0.43
Schools in Suburban Setting	0 = no 1 = yes	0.53	0.50
Grade 8 Enrollment (in hundreds)	(0.01, 16.07)	2.17	1.39
<u>State-level Data</u> (42 states)			
Percent of Children in Poverty	(7, 33.50)	17.72	5.92
Educational Expenditure per Pupil (in hundreds)	(25.47, 78.27)	45.22	12.35

4.3 Results

4.3.1 Unconditional Model

Table 4.2 gives the posterior characteristics of the marginal distributions of the parameters in the unconditional model. Listed are the means, the standard deviations of the intercept or the grand mean log odds of offering algebra and between-state variance, the 90% credibility interval (C.I.) for each parameter, and the mode for the variance estimate. These statistics were computed from 1000 simulated values obtained after convergence. The results show that on average the logit of offering algebra is 1.312 and there is statistically significant state-to-state variation in the log-odds of offering algebra. The predicted odds ratio is $exp\{1.312\} = 3.714$. Schools are more likely to offer the course than not on average. The predicted probability is $(1+exp\{-1.312\})^{-1} = 0.789$. Note that this differs from the mean outcome across the whole sample, which is equal to 0.76. The difference arises because the grand mean estimated here is based on the conditional distribution of the outcome given the random effects are null, rather than on the marginal distribution of the outcome (Zeger, Liang & Albert, 1988; Raudenbush, 1993).

Table 4.2: Results of the Unconditional (No Predictors) Model

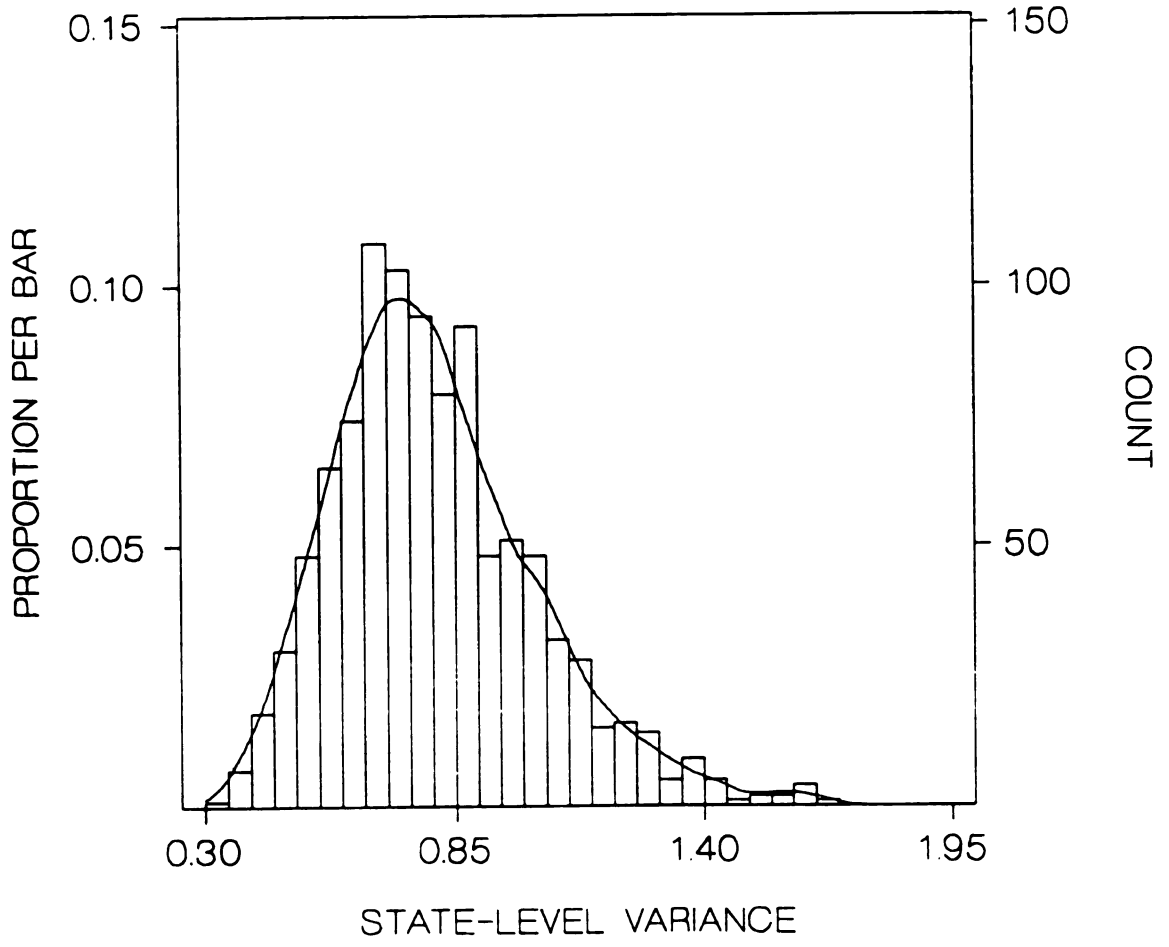
Explanatory Variables	Posterior Characteristics			
	Mean	Std. Dev.	90% C.I.	
			From	To
Intercept				
Intercept	1.312	0.145	0.886	1.810
State-to-state Heterogeneity				
τ (mode = 0.730)	0.806	0.223	0.492	1.222

Figure 4.3, a histogram that approximates the posterior distribution of the between-state variance of the log odds based on 1000 sampled values of τ, portrays the variability. As Figure 4.3 implies, values of τ as small as 0.30 and as large as 1.70 are plausible, whereas the 90 C.I. covers from 0.49 to 1.22. A sense of the magnitude of state-to-state differences in the likelihood of offering algebra can be conveyed by computing a plausible range of predicted probabilities with the random effects varying from -1.5 to 1.5 standard deviation units, i.e., $p_{ij} = 1/(1+\exp\{-\beta_{oj}\})$ for

$$\beta_{0j} \in (\gamma \pm 1.5\sqrt{\tau}) ,$$

where β_{oj} is the average log odds of offering algebra for schools in state *j* (Raudenbush, 1993). When τ is equal to the mode of the posterior, 0.806, the predicted probabilities range from 0.51 to 0.93; when τ is at the 10th percentile, 0.48, they range from 0.57 to 0.91; and when τ is at the 90th percentile, the range is from 0.41 to 0.95. The ranges indicate substantial state differences in the predicted probabilities of offering algebra.

Figure 4.3: Marginal Posterior Distribution of τ (The Unconditional Model)



For individual states, Figures 4.4a and 4.4b display box plots describing the distribution of the predicted probabilities computed using the marginal posterior of $\boldsymbol{u}$. Each box plot summarizes 1,000 probabilities computed from 1,000 corresponding simulated values of mean logits. Inside the box are the middle 75 percent of the values of the probabilities, while the ends of the whiskers delimit the top and bottom 1 percent of those values. Thus, the plots give 75 percent and 98 percent intervals for the predicted probabilities of offering algebra for each state. The figures offer another graphical assessment of between-state variability. For instance, there are quite a few non-overlapping distributions. PA, for example, is shown to differ from OK and TN in its predicted probabilities of offering algebra.

Figure 4.4a

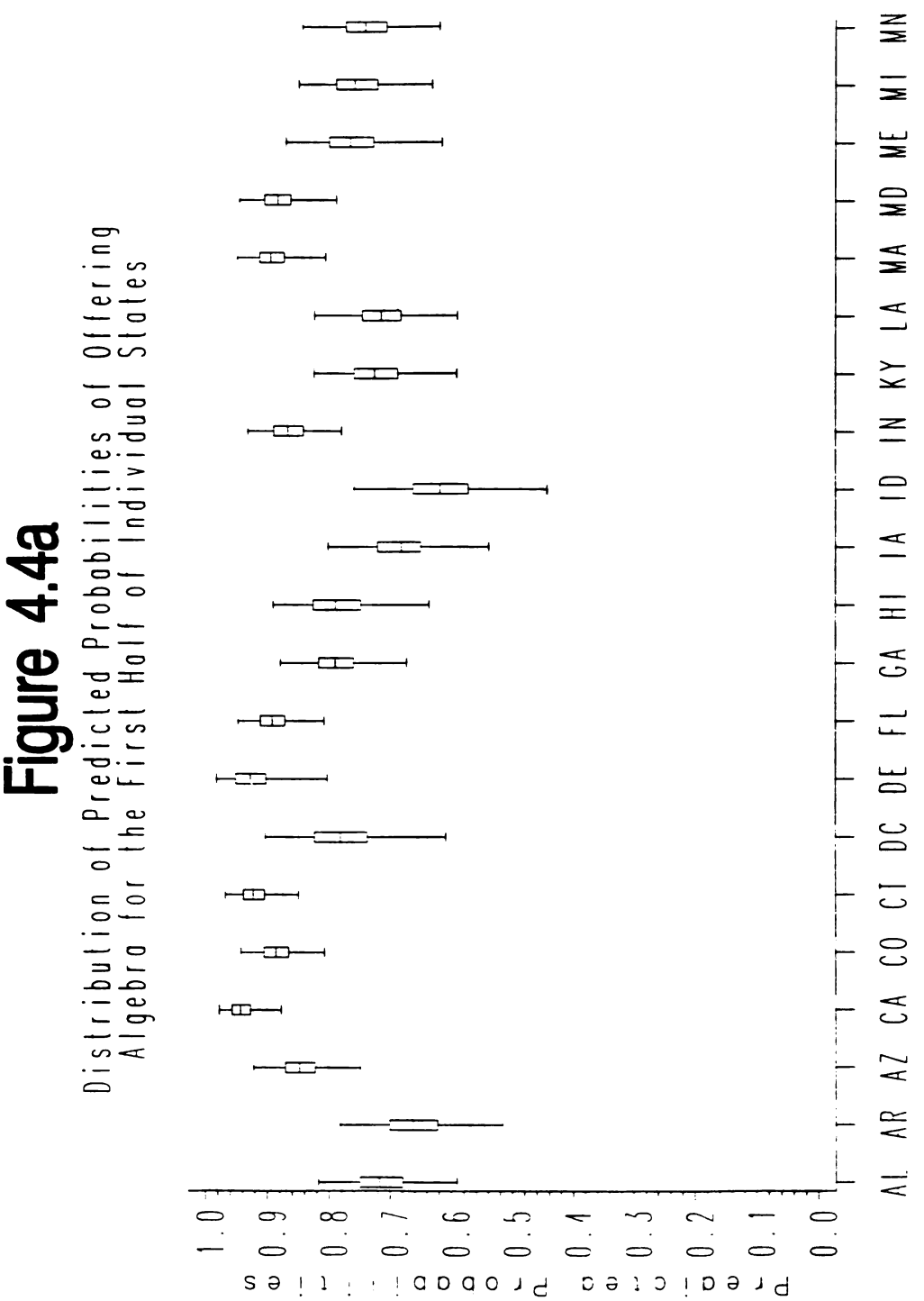
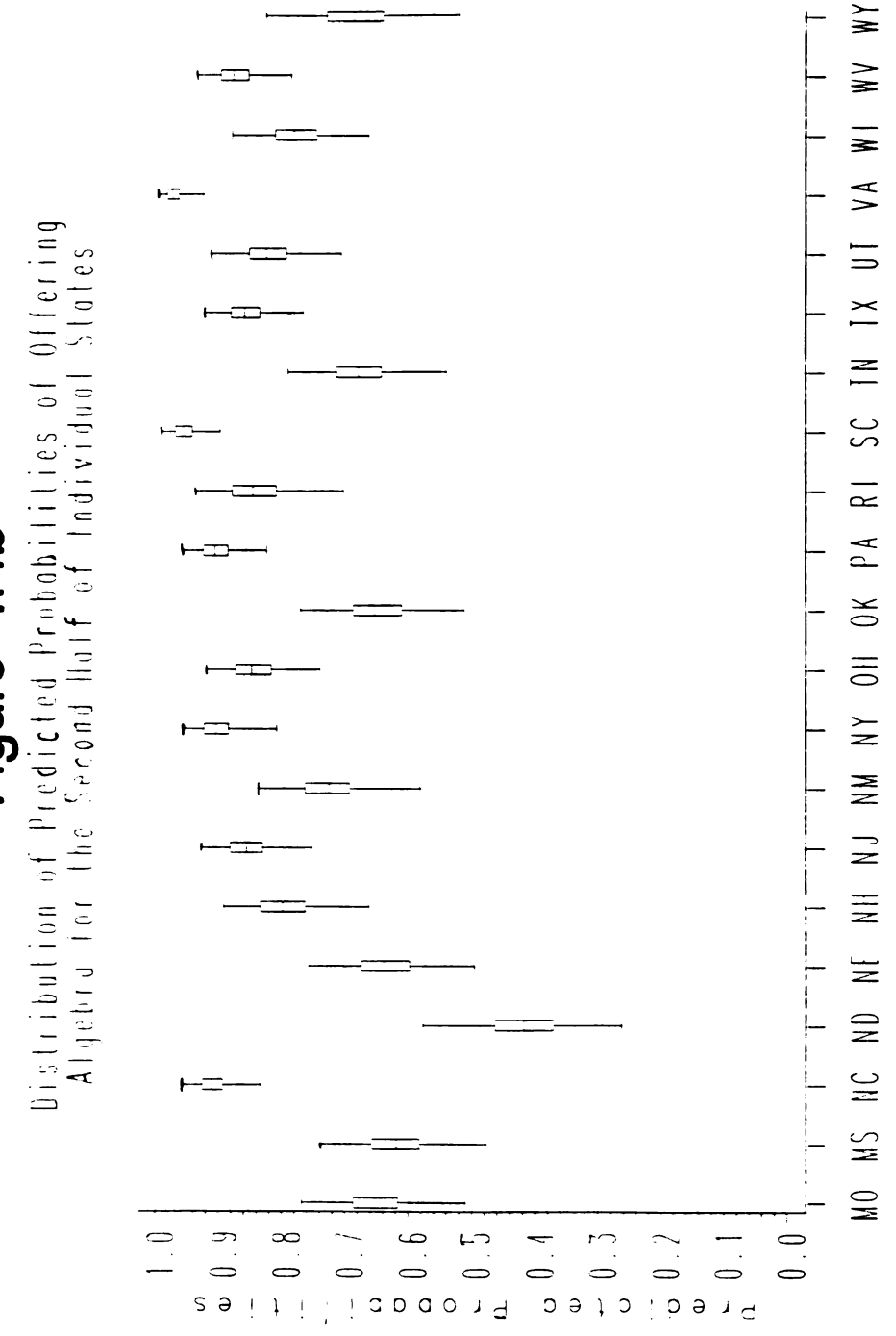

Distribution of Predicted Probabilities of Offering
Algebra for the First Half of Individual States
Predicted Probabilities
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
AL AR AZ CA CO CT DC DE FL GA HI IA ID IN KY LA MA MD ME MI MN
First Half of States Displayed

Figure 4.4b


Distribution of Predicted Probabilities of Offering
Algebra for the Second Half of Individual States
Predicted Probabilities
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
MO MS NC ND NE NH NJ NM NY OH OK PA RI SC TN TX UT VA WI WV WY
Second Half of States Displayed

4.3.2 Between-school/Within-state and Between-State Model

The second model consisted of between-school/within-state and between-state predictors. Table 4.3 summarizes the result and displays posterior characteristics of the marginal distributions of the parameters, based on successive sets of 1000 sampled values collected after convergence, for the school- and state-level predictors and the state-level variance.

Table 4.3: Results of the Within- and Between-State Model

Explanatory Variables	Posterior Characteristics			
	Mean	Std. Dev.	90% C.I.	
			From	To
Intercept				
Intercept	1.540	0.121	1.349	1.747
State Expenditure Per Pupil	0.044	0.012	0.023	0.063
Percent of Student in Poverty	0.016	0.023	-0.018	0.053
Minority Enrollment				
Intercept	-0.664	0.161	-0.939	-0.401
State Expenditure Per Pupil	-0.046	0.013	-0.067	-0.024
SES				
Intercept	0.383	0.068	0.267	0.496
State Expenditure Per Pupil	0.008	0.006	-0.001	0.017
Urban				
Intercept	0.398	0.159	0.131	0.651
Suburban				
Intercept	0.230	0.117	0.105	0.486
Grade 8 Enrollment				
Intercept	0.645	0.050	0.562	0.725
Grade 8 Enrollment (Quadratic Term)				
Intercept	-0.073	0.013	-0.097	-0.053
State-to-state Heterogeneity				
τ (mode = 0.437)	0.477	0.147	0.286	0.736

At the school level, the log odds of offering algebra is negatively related to percent of minority enrollment and positively related to school SES, urban and suburban setting, and size. A sense of the magnitude of the effects of specific predictors can be conveyed by computing the scale-invariant measure of odds ratio, $\exp\{\gamma\}$, which yields the ratio of predicted odds of offering algebra, or by obtaining the predicted probabilities of the existence of the course for schools which differ with respect to particular characteristics.

I first examine the effects of the percent of minority enrollment on the likelihood of offering algebra. The estimated ratio of the odds of offering algebra for schools which have minority enrollment greater than or equal to 50% versus schools which do not is $exp(-0.664) = 0.515$, after controlling for the effects of other variables in the model. It indicates on average that the odds of offering the course among typical schools with high minority enrollment are about half of those with less minority enrollment.

Another way to convey the effect of minority enrollment is to compute the predicted probabilities for two typical schools which differ in their racial composition. The probabilities can be obtained by holding all of the grand-mean centered predictors constant at zero and allowing only the minority enrollment indicator to vary. As Table 4.1 shows, the mean of high minority enrollment is 0.14, indicating that 14% of the sample had minority enrollment in excess of 50%. The grand-mean centered version of minority enrollment takes on a value of 1 - 0.14 = 0.86 for schools with high minority enrollment and 0 - 0.14 = -0.14 otherwise. Given random state effects of zero ($u_{0j} = 0$), the predicted probabilities of offering algebra are $(1 + exp\{-[1.540 + -0.664 * 0.86]\})^{-1} = 0.725$ for a

typical school with high minority enrollment and $(1 + exp\{-[1.540 + -0.664 * -0.14]\})^{-1} = 0.837$ for a typical school with fewer minority students, other things being equal.

As Table 4.3 implies, school SES composition is associated with the log odds of offering algebra, net of the effects of other variables. The likelihood of offering the advanced mathematics course is higher among schools with higher SES. To gauge the effect, the odds ratio for two typical schools which are two standard deviations apart in their SES composition (2*0.82) = 1.64 is computed. The odds ratio estimated for the higher SES versus lower SES schools is 1.874 *(exp*(0.383*1.64)). Given random state effects are null, the predicted probabilities for the two typical schools to offer algebra are $(1 + exp\{-[1.540 + 0.383 * 0.82]\})^{-1} = 0.865$ and $(1 + exp\{-(1 + exp\{-[1.540 - 0.383 * -0.82]\})^{-1} = 0.773$.

Rural schools, when compared with their urban and suburban counterparts, are in general less likely to include 8th grade algebra in their mathematics curricula, when controlling for the school demographic composition, size, state educational expenditure, and percent of children in poverty. The estimated ratios of odds for an urban versus a rural school and a suburban versus a rural school in offering the advanced mathematics course are 1.488 and 1.259 respectively. The predicted probabilities for typical urban, suburban, and rural schools to offer algebra are 0.864 $((1 + exp\{-[1.540 + 0.398 * 0.77]\})^{-1})$, 0.839 $((1 + exp\{-[1.540 + 0.23 * 0.47]\})^{-1})$, and 0.784 $((1 + exp\{-[1.502 + 0.398 * -0.23 + 0.23*-0.53]\})^{-1})$.

The partial effect of school size on the logits of offering algebra, net of those of all other predictors, is positive and statistically significant. The relative odds for schools with

eighth-grade enrollments of 356 (one standard deviation above the average) to schools with average eighth-grade enrollments are $exp(0.645*1.39 + -0.073*1.39*1.39) = 2.129$, indicating the odds that the former schools offer algebra are twice as likely as the latter. The predicted probability of offering algebra for a typical school with 356 eighth-grade students is 0.909, whereas it is 0.824 for a typical one with average eighth-grade enrollment. Note that 0.824 here differs from 0.789 estimated in the unconditional model as it is conditioned on the random effect after controlling for state educational expenditure and percent of children in poverty.

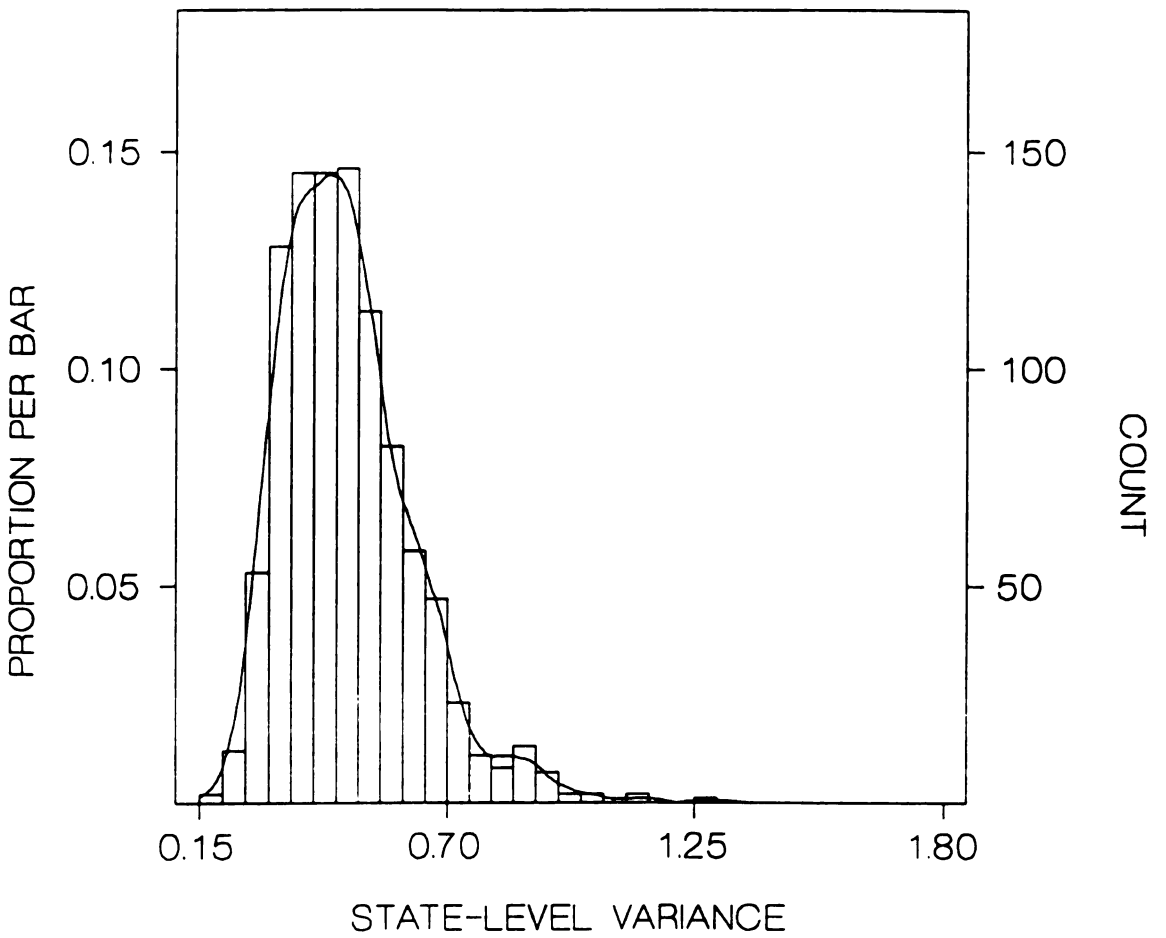
At the state level, the log odds of offering algebra, adjusted for school composition, urbanicity, and size, are associated with state educational expenditure per pupil but not percent of children in poverty. The estimated odds ratio for schools in states with educational expenditure per pupil one standard deviation above the average versus their counterparts in states with average educational expenditure is $exp(.044*12.35) = 1.72$, after controlling for the percent of student in poverty and other school-level predictors. The predicted probabilities of offering algebra for two typical schools, one located in a state with an educational expenditure one standard deviation above average, the other in a state with average educational expenditure, are 0.824 and 0.889, respectively.

The relationship between minority enrollment and the likelihood of offering algebra depends on state educational expenditure per pupil. The gap in the log odds of offering algebra between predominantly minority schools and schools with fewer minority students is widened as state educational expenditure per pupil increases. The

estimated odds ratio for high minority schools in states with educational expenditure per pupil one standard deviation above the average versus their counterparts in states with average educational expenditure is $exp(-.046*12.35) = 0.568$. The predicted probabilities of offering algebra for two typical schools located in a state with an educational expenditure one standard deviation above average, with and without high minority enrollment, are 0.670 and 0.725 respectively.

As the result in Table 4.3 suggests, the relationship between school SES and how likely a school is to offer algebra does not depend on state per pupil educational expenditures. Regarding the state-to-state heterogeneity, substantial variance remains to be accounted for after putting the school- and state-level predictors into the model. There is about a reduction of 35.6% in the mode of the variance. Figure 4.5 gives a histogram of the posterior distribution of the conditional variance for the second model. Note that the 90% C.I. (0.286, 0.736) of the adjusted or residual variance is much narrower than that of the unconditional model 90% C.I. (0.492, 1.222). Allowing the random effect to vary from -1.5 to 1.5 standard deviation units and holding all the variables constant at their grand means, the adjusted predicted probabilities for individual states range from 0.63 to 0.93 when τ is equal to its mode, 0.437.

Figure 4.5: Marginal Posterior Distribution of τ (The Conditional Model)



4.4 Discussion and Implications

Given that the results of the analyses confirmed most of the hypotheses stated in the study, a question that arises is whether or not the statistically significant partial effects of the school- and state-level predictors are practically significant as well. Translating the changes in odds or predicted probabilities to the opportunities to learn at the student level allows us to assess the practical significance and the magnitudes of effects of the predictors. It enables us to examine the extent to which factors at one level of the multilevel educational system influence, alter, or limit the events at lower levels.

To proceed, I use the predicted probabilities of offering algebra obtained for two typical schools with and without high minority enrollment. All else held constant, the two predicted probabilities are 0.725 and 0.837, respectively. They differ by 0.112. Thus, for 100 typical schools with high minority enrollment, it is predicted that there will be 11 fewer schools offering algebra than if the probabilities were equal. As seen in Table 4.1, the average grade eight enrollment is 217. Approximately 2,400 students will attend schools that do not offer algebra for high school credits as a result. This indicates that the practical significance of the racial composition effect is substantial. By the same token, differences in the predicted probabilities of offering algebra among typical schools with different SES composition, size, and in different settings (rural versus others) could have a significant impact on students' access to the course.

The correlational effect of the denial of access to algebra initiated at a higher level of the educational system can also be illustrated using a unit above schools--the state. For example, all else being equal, a difference of 0.065 in the predicted probabilities of

offering algebra brought about by a one standard deviation decrease in state educational expenditure per student will be translated to the absence of the algebra course in hundreds of typical schools within a state and, on average, to the preclusion of possibilities to learn algebra for 217 students in each school.

Whereas the partial effects allow us to gauge the influence of a particular factor net those of the other factors, it may be useful and even important to examine how the outcome varies with more than one variable. One reason is that a school's disadvantage in its capacity to offer algebra may be compounded by other factors. For instance, high minority schools on average have lower SES than low minority schools, and rural schools usually have fewer students than their urban and suburban counterparts. In the 1992 TSAP sample, high minority schools have a mean SES of -0.982, and low minority schools have a mean of 0.163. The two means are more than one standard deviation apart. Thus, other factors remaining equal, the predicted probability of offering the course for a typical high minority school whose SES is equal to -0.982 is (1+exp{-[1.540 + -0.664*0.86 + 0.383*-0.982]}) = 0.644. It is 0.081 less than that for a high minority school with average SES, i.e., 0.

After assessing the practical significance of most predictors, I now offer and discuss some tentative explanations as well as the implications of the findings. The analyses indicate racial and social stratification in the opportunities to learn algebra among schools in the 42 states examined. The fact that high minority schools are more likely to have lower SES suggests there is a clustering of poor and minority children in schools, which leads to a compounding of school's disadvantage. The clustering, as

suggested by Hawley (1993) and Carter (1995), could have been brought about by residential segregation, policies of neighborhood school assignment, and school district configurations.

Regarding school size, the results support the assertion that greater grade enrollment encourages greater specialization and differentiation of the curriculum (Lee & Smith, 1995). The association, however, is a nonlinear one and it weakens as school size increases. One possible reason that rural schools are less likely to offer the advanced mathematics course, among other things, may be the limited availability of highly-qualified teachers in the rural sector due to the more restrictive markets in certain demographic regions (Monk & Haller, 1993). There may be an inequitable distribution of well-trained teachers across the different sectors.

The association of increased levels of financial investment in public education with a greater access to the course lend some support to the conclusions of Hedges, Laine, and Greenwald's (1994) meta-analysis of studies of the effects of differential school inputs on student outcomes. They concluded that financial resource input is likely to be related to school outcomes. The result of the present investigation shows that increased state spending is associated with the offering of a course and in turn is associated with higher mathematics achievement (MacIver & Epstein, 1995) and a greater likelihood of enrollment in advanced high school mathematics (Stevenson et al., 1984). It is important to note here, however, as no educational expenditure at the district level was entered as a predictor, district-to-district variability in funding was not modeled. In order to model the relationship between the two educational inputs (district educational spending and the

availability of the course), two indices of educational investment efforts would be needed. These indices are the educational expenditure per student as well as the level of special education funding. TSAP does not provide the relevant data for the inclusion of these indices in the analysis. Hence what was examined here was the general effect of overall state public education investment efforts on the availability of the advanced mathematics course.

Whereas higher levels of public education investment efforts in public education are associated with a greater likelihood of the offering of the course, they also widen the racial composition gap in access to algebra. This seems to suggest that increased state educational expenditures benefit low-minority schools more than high-minority schools.

In sum, the results suggest that schools serving minority students and low SES students, small schools, and schools located in states spending less on education are comparatively less likely than other schools to offer algebra. The implication of the finding is that size, composition, and location of a school are linked to inequality in access to this particular educational resource. In these ways, the schooling system reinforces social, ethnic, and geographic inequality in the opportunity to study high-school algebra in the middle grades.

Finally, the analyses show that in addition to allowing one to analyze the specific effects of variables of interest, while holding other factors constant, the fitted or predicted probabilities could act as a production and curriculum indicator (Oakes, 1989; Pelgrum et al., 1995; Porter, 1991) at the state and the school level. They inform us about the "implemented curriculum" (Pelgrum et al., 1995) at those levels and how they are related

to "curriculum antecedents" (Pelgrum et al., 1995) such as the SES composition of the school. The fitted probabilities could perform as an indicator as well as a diagnostic of how well the educational systems perform. For example, they can indicate the presence of the stratification of learning opportunities due to race, SES, and location in terms of student attendance at a school that offers algebra.

CHAPTER 5

SUMMARY AND CONCLUSIONS

5.1 Summary and Conclusions

In this work, I coded Zeger and Karim's (1991) Bayesian posterior analysis for generalized linear models with random effects via the Gibbs sampler. The SAS/IML code performed well and the algorithm gave reasonable inferences. This was documented with the help of a simulated study with datasets having structure similar to the TSAP in mathematics data. Using the algorithm, I implemented a Bayesian, multilevel analysis of access to eighth-grade algebra. The analysis accommodated the hierarchical or nested design of the TSAP in mathematics and the need to incorporate the uncertainty that arose about the variances at the state level.

The results suggest that schools serving minority students and low SES students, small schools, schools located in the rural setting, and states spending less on education are comparatively less likely than other schools to offer algebra. The implication is that size, composition, and location of a school are linked to inequality in access to this educational resource. In these ways, the schooling system reinforces social, ethnic and geographic inequality in the opportunity to study advanced mathematics. The analyses demonstrated that the fitted probabilities could be used as an indicator for certain academic opportunities in schools with different demographics, enrollment, setting, and levels of state financial investment efforts. In addition, the approach can be applied to policy research addressing the multilevel structure of the educational and social systems.

5.2 Future Research Needs

5.2.1 Substantive Inquiry

This study provided a snapshot of the different availability of eighth-grade algebra among public schools in 42 states in 1992-93. The focus of the inquiry is on an essential component of the opportunities to learn algebra in middle grades, which is content of instruction that schools make accessible to students (Porter, 1994). With the current reforms in mathematics that stress a high-quality curriculum for all students (Porter, 1994; Wheelock, 1996), efforts to effect changes in the content of and access to algebra in the mathematics curriculum for the middle grades have been initiated over the past few years (Edwards, 1994; Olson, 1994).

Two main and related curricular changes regarding algebra instruction are advocated by the widely-circulated Curriculum and Evaluation Standards for School Mathematics (NCTM, 1989), often abbreviated as the Standards. First, as Jack Price, the president of NCTM, stated, algebra should be seen as "a strand throughout the K-12 curriculum" (Olson, 1994, p.11), and not as an isolated instructional topic. Indeed it is recommended not to segregate basic algebraic ideas into a one-year course (NCTM, 1993). Second, algebraic thinking that emphasizes understanding, reasoning, problem-solving, and real-world applications is to be introduced gradually to all students in a broad and integrated curriculum for the middle grades. Other topics included in the recommended curriculum are measurement, geometry, probability and statistics. An example of an integrated middle grades mathematics curriculum that reflects the Standards is the Connected Mathematics Project Curriculum (Connected Mathematics

Project, 1996).

According to the proponents of the Standards, extensive curriculum development is required to guarantee access to algebraic competence for all students (Silver, 1995). For supporters, the traditional first-year course in algebra is not the "right algebra" for everyone (Chambers, 1994), in part because of the course puts undue emphasis on rules and algorithms.

The requirement of eighth-grade algebra in various school districts such as Cambridge, Massachusetts (Olson, 1994) has brought greater access to algebra to students. The district requirement shares the goals of the Standards to engage more students in learning important mathematical ideas and promote their proficiency in mathematics (Silver, 1995).

Further research can be carried out to study the impacts of current reforms on access to algebra. Of particular interest is to examine the availability of algebra instruction in the middle grades amidst the perceived incompatibility between the advocacy of a "problem-based" curriculum suggested by the NCTM and the more conventional "scope-and-sequence" approach (Schoenfeld, 1994). In addition, the great amount of instructional and financial support and community involvement called for by the standard-based reforms, for example, extensive curriculum development and teacher education, add additional difficulties in the reform process to guarantee access and opportunity for all students (Bruckerhoff, 1995; Silver, 1996; Tate, 1994). An important research question one may ask is what state- and school-level factors may influence a school's decision to implement curricular reform and offer "problem-based" algebra

instruction in the middle grades.

The present inquiry focused on the existence of an algebra course for high school credits in the middle grades. Further research can be carried out to review its "breadth" or the percentage of students within a school who enroll in the course. Becker (1990) found substantial within-school variability in access to algebra. He found that out of the 63% of the 2,400 middle-grade schools which offered algebra courses in their curriculums in grade 7, and, more typically, grade 8, only 35% provided access to algebra to at least one-quarter of their students. A future study of the student- and school-level correlates of enrollment in the algebra course can complement the present investigation by offering an understanding of the selection process at another level of the hierarchical educational system, that of the student. Using Pelgrum et al.'s (1995) term, the study will allow one to study the "attained curriculum" (p. 81) at the micro, or student level, and will complement the present investigation of the "implemented curriculum" (p.81) at the meso, or school and state levels, in monitoring the performance of the multilevel educational system.

5.2.2 Methodological Development

The Gibbs sampler was coded in SAS/IML and its execution was very time consuming. In the simulation, for instance, it took approximately 8 to 16 hours to fit a model. Coding in a lower or intermediate-level language such as C will greatly reduce the run times. A prototype written in C is currently being developed. My early prediction is that it will take approximately half an hour to forty five minutes to execute the same analysis mentioned previously.

The simulations and analyses of the present study focused on cases where τ is a scalar. With the development of a faster computer code, simulations to document the performance of the algorithm for cases where T is a matrix can be carried out. Satisfactory results will allow the code to be applied to analyze random coefficient models. A slight modification of the over-simplified model **(2.6)** described in Chapter 2 gives us an example of a random coefficient model. By assuming the racial composition effect to be randomly varying across the 42 states after controlling for $StateExp_j$, i.e.,

$$\beta_{1j} = \gamma_{10} + \gamma_{11} StateExp_j + u_{1j} , \qquad \textbf{(5.1)}$$

model **(2.6)** becomes a random coefficient model with $\boldsymbol{u}_j$ being a 2 x 1 vector of random effects with mean of zero and variance T.

APPENDICES

APPENDIX A

APPENDIX A

SAS\IML Code for the Gibbs Sampler

```
/*************************************************************************/
/*

Program name:  bayes.sas

Purpose and Modules:  The code implements the Gibbs sampler developed
by Karim (1991) and Zeger and Karim (1991). Its main program consists of
the following modules:

        1) import -- reads in the data set,
        2) initial -- enters various parameter values and sets up vectors and
                      matrices,
        2) glimgam -- performs GLIM to obtain γ̂ and V_γ
        3) gbgam -- generates γ
        4) gbtau -- generates T
        5) ridgeuj -- performs ridged regression to obtain û_j and V_ûj
        6) gbbj -- generates u_j
        7) output -- outputs simulated values of γ, T and u_j
*/
/*************************************************************************/

/******** formats and directories ******/

Options pagesize= 100 linesize = 80 nocenter number date;

Libname gibbs 'c:\sas\data';
Libname output  'c:\sas\output';

Proc iml;
reset nolog;

/********* formats and directories ******/
```

```
/********* module import  ************/

start import (yij,lev2var,lev1var,nj);  /* begin module import: get raw data */

/* yij          = outcome variable
 lev2var       = a level 2 variable
 lev1var       = a level 1 variable
 nj            = number of level 1 units in cluster j    */

use gibbs.rawdata;
read all var{yij lev2var lev1var};
close metgibbs.rawdata;

use gibbs.nj;
read all var {nx} into nj;
close metgibbs.nj;

finish;

/***** module import ********/
```

```
/*********Part 2 ************/

start initial (yij,lev2var,lev1var,nj,xij,zij,ngroup,N,nf,nq,inititer,
  niter,nsample,gam,uj,Tau,seed1,outgam,outuj,outTau);

/* xij          = predictors for fixed effects
 zij           = predictors for random effects
 N             = total sample size
 ngroup        = number of level-2 units
 nf            = number of fixed effects --- beta's
 nq            = number of random effects --- z's
 inititer      = counter of iteration
 niter         = number of maximum iteration
 nsample       = sample size of the posterior dist.
 gam           = fixed effects
 uj            = random effects
 Tau           = variance of uj's
 seed1         = seed1 number
 outgam        = output beta 0 = no, 1 = yes
 outuj         = output bj   0 = no, 1 = yes
 outTau        = output Tau    0 = no, 1 = yes */

ngroup         = 40;
nf             = 3;
nq             = 1;
inititer       = 10;
niter          = 5000;
nsample        = 2000;
seed1          = 12378942;
outgam         = 1;
outbj          = 1;
outTau         = 1;

bj = j(ngroup*nq,1,0);
yij = yij;
N = nrow(yij);
x1 = j(N,1,1);
xij = x1||lev2var||lev1var;
z1 = j(N,nq,1);
zij = z1;
finish;

/******* module initial *******/
```

```
/******** module glimgam  ********/

start glimgam (yij,nj,nf,nq,ngroup,N,xij,zij,gam,uj,vgam,inititer,det,
            loggamx,ofstuj);

crit=1;
lo=0;
iterval=j(20,2,0);
loggamx = 0;
ofstuj = j(N,1,0);          /* offset zijuj              */

a = 1;
c = 1;
e = 1;
f = 1;
g = nq;

Do i = 1 to ngroup;
      ijth = a - 1 + nj[c:e];
      ofstuj[a:ijth] = zij[a:ijth,1:nq]*uj[f:g];
      a = a + nj[c:e];
      c = c + 1;
      e = e + 1;
      f = f + nq;
      g = g + nq;
End;

Do it=1 to 20 while(crit> 1.0e-8);
etaij = j(N,1,0);
loglhood=0;
xwx = j(nf,nf,0);
xwystar = j(nf,1,0);
ystarij = j(N,1,0);
wij = j(N,1,0);
uij = j(N,1,0);
a = 1;
c = 1;
e = 1;
f = 1;
g = nq;
```

```
Do i = 1 to ngroup;
      ijth = a - 1 + nj[c:e];
      etaij[a:ijth] = xij[a:ijth,1:nf]*gam + ofstuj[a:ijth];
            nunit = ijth - a + 1;
      if inititer <= 10 then
      do;
      if it <= 3  then
      do;
      h = a;
      do k = 1 to nunit;
      if  .95 <= uij[h:h]  then
            do;
                  uij[h:h] = .95;
                  * print 'large probability';
            end;
            h = h+1;
      end;
      end;
      end;
      wij[a:ijth] = uij[a:ijth]#(1-uij[a:ijth]);
      ystarij[a:ijth] = etaij[a:ijth] +
            ((yij[a:ijth] - uij[a:ijth])/wij[a:ijth]);
      xwx  = xwx + (xij[a:ijth,1:nf]#wij[a:ijth])`*xij[a:ijth,1:nf];
      xwystar = xwystar + ((xij[a:ijth,1:nf]#wij[a:ijth])`*(ystarij[a:ijth]-
            ofstuj[a:ijth]));
      loglhood = loglhood + sum(yij[a:ijth]#log(uij[a:ijth]) +
            (1-yij[a:ijth])#log(1-uij[a:ijth]));

      a = a + nj[c:e];
      c = c + 1;
      e = e + 1;
      f = f + nq;
      g = g + nq;
end;

gam = solve(xwx,xwystar);
vgam = inv(xwx);
crit=abs(loglhood-lo);
iterval[it,]=it||loglhood;
lo = loglhood;

end;
```

```
loggamx = loglhood;

finish;

/***** module gimgam ****/
```

```
/********* module gbgam ******/

start gbgam (yij,nj,nf,nq,ngroup,N,xij,zij,gam,uj,vgam,seed1,loggamx,ofstuj,inititer);

/* ygam      = envelope function (multivariate normal distribution)
xgam        = actual function
c1          = mode-matching constant
c2          = variance-inflation constant
loggamx     = pi(x) from glimgam */

ygam = 0;
xgam = 0;
gamstar1 = j(nf,1,0);
c1 = 0;
c2 = 2;
etaij = j(N,1,0);
uij = j(N,1,0);
wij = j(N,1,0);
nordev = j(nf,1,0);
ratio = 0;

/* compute the log of ordinate of ygam--the envelope function at the mode */

loggamy = - ((nf/2) * log (c2 * 3.1416))
        - ((nf/2) * log (det(c2 * vgam)));


/* compute c1 */

c1 = loggamx - loggamy;
print c1;

/* generate gamstar from ygam */

u = 1;
gamcnt = 0;
Do while (ratio < u);

a = 1;
b = 1;
do i = 1 to nf;
      nordev[a:b] = rannor(seed1);
      a = a + 1;
```

```
      b = b + 1;
end;

CholeskU = j(nf,nf,0);
CholeskU = root(c2*vgam);
gamstar= gam+CholeskU`*nordev;

/* compute the ordinate */

loggamy = - ((nf/2) * log (c2 * 3.17))
        - ((nf/2) * log (det(c2 * vgam)))
        - ((gamstar-gam)`*inv(c2*vgam)*(gamstar-gam))/2;

/* compute the ordinate of xgam */

a = 1;
c = 1;
e = 1;
f = 1;
g = nq;
loggamx = 0;

Do i = 1 to ngroup;
      ijth = a - 1 + nj[c:e];
      nijth = 0;
      nijth = ijth - a + 1;
      etaij[a:ijth] = xij[a:ijth,1:nf]*gamstar + ofstuj[a:ijth];
      uij[a:ijth] = 1/(1+exp(-1*etaij[a:ijth]));
      loggamx = loggamx + sum(yij[a:ijth]#log(uij[a:ijth]) +
            (1-yij[a:ijth])#log(1-uij[a:ijth]));
      a = a + nj[c:e];
      c = c + 1;
      e = e + 1;
      f = f + nq;
      g = g + nq;
end;

/* calculate the ratio fgam/ggam */

gamcnt = gamcnt + 1;

ratio = exp(loggamx - loggamy - c1);
u = uniform(seed1);
```

```
end;

/* updating gam */

gam = gamstar;
end;

finish;

/******* module gbgam *********/
```

```
/******* module gbTau ***************/

start gbTau (uj,Tau,nq,ngroup);

seed1 = 344456;
S = j(nq,nq,0);

c = 1;
e = nq;

Do i = 1 to ngroup;
S = S + (uj[c:e]*uj[c:e]`);
c = c + nq;
e = e + nq;
end;

do i = 1 to nq;
      do j = 1 to nq;
            if (i < j) then S[i,j] = S[j,i];
      end;
end;

invS = j(nq,nq,0);

invS = inv(S);

do i = 1 to nq;
      do j = 1 to nq;
            if (i < j) then invS[i,j] = invS[j,i];
      end;
end;


H = root(invS);
W = j(nq,nq,0);
do i = 1 to nq;
  do j = 1 to nq;
    if (i < j) then do;
       W[i,j] = rannor(seed1);
    end;
    if (i = j) then do;
       df = ngroup - nq - 1;
       * df = ngroup - j + 1;
```

```
      if (mod(df,2) = 0) then
      do;
            df = int(df/2);
            W[i,j] = sqrt(2.0 * rangam(seed1,df));
      end;
      else do;
            df = int(df/2);
            W[i,j] = sqrt(2.0 * rangam(seed1,df) + rannor(seed1)**2);
      end;
   end;
 end;
end;


Tau = inv(H` * W` * W *  H);

do i = 1 to nq;
      do j = 1 to nq;
            if (i < j) then Tau[i,j] = Tau[j,i];
      end;
end;


*

finish;

/******* module gbTau **************/
```

```
/******  module ridgeuj **************/

start ridgeuj  (yij,xij,zij,nj,N,ngroup,nf,nq,gam,Tau,uj,varuj,inititer,
            logujx,ofstgam);

crit=1;
lo=0;
iterval=j(20,2,0);

ofstgam = j(N,1,0);          /* offset xijgam              */

a = 1;
c = 1;
e = 1;
ijth = 0;

Do i = 1 to ngroup;
ijth = a - 1 + nj[c:e];
ofstgam[a:ijth] = xij[a:ijth,1:nf]*gam;
a = a + nj[c:e];
c = c + 1;
e = e + 1;
End;


ystarij = j(N,1,0);
wij = j(N,1,0);
uij = j(N,1,0);
etaij = j(N,1,0);
varuj = j(ngroup*nq,nq,0);
a = 1;
c = 1;
e = 1;
f = 1;
g = nq;
ijth = 0;
loglhdp1 = j(N,1,0);
logujx = j(ngroup,1,0);
```

```
Do i = 1 to ngroup;

ijth = a - 1 + nj[c:e];
nijth = ijth - a + 1;
zwzinvT = j(nq,nq,0);
zwystaro = j(nq,1,0);
crit = 1;
it = 0;
lo = 0;

      Do it = 1 to 20 while (crit> 1.0e-8);
      loglhood = 0;
      etaij[a:ijth] = ofstgam[a:ijth]+
            + zij[a:ijth,1:nq]*uj[f:g];
      uij[a:ijth] = 1/(1+exp(-etaij[a:ijth]));
      nunit = ijth - a + 1;
      if inititer <= 50 then
      do;
      if it <= 20  then
      do;
      h = a;
      do k = 1 to nunit;
      if  uij[h:h] <= .01 then
            do;
                  uij[h:h] = .01;
            end;
      else if .99 <= uij[h:h] then
            do;
                  uij[h:h] = .99;
            end;
            h =h+1;
      end;
      end;
      end;
      wij[a:ijth] = uij[a:ijth]#(1-uij[a:ijth]);

      ystarij[a:ijth] = etaij[a:ijth] +
            (yij[a:ijth] - uij[a:ijth])/wij[a:ijth];
      zwzinvT = (zij[a:ijth,1:nq]#wij[a:ijth])`*zij[a:ijth,1:nq]+inv(Tau);
      zwystaro = (zij[a:ijth,1:nq]#wij[a:ijth])`*(ystarij[a:ijth]-
            ofstgam[a:ijth]);

      loglhdp1[a:ijth] = (yij[a:ijth]#log(uij[a:ijth]))
```

```
            + ((1-yij[a:ijth])#log(1-uij[a:ijth]));
        loglhood = sum(loglhdp1[a:ijth]) - (uj[f:g]`*inv(D)*uj[f:g])/2;
           crit = abs(loglhood-lo);
        iterval[it,]=it||loglhood;
        lo = loglhood;
        uj[f:g] = solve(zwzinvT, zwystaro);
        end;

varuj[f:g,1:nq]= inv(zwzinvT);

logujx[c:e] = loglhood;

a = a + nj[c:e];
c = c + 1;
e = e + 1;
f = f + nq;
g = g + nq;
end;


finish;

/******* module ridgeuj **********/
```

```
/******** module gbuj ************/

start gbuj  (yij,xij,zij,nj,N,ngroup,nf,nq,gam,Tau,uj,varuj,seed1,
            logujx,accuj,ofstgam,inititer);


CholeskU = j(ngroup*nq,nq,0);
u = j(ngroup*nq,1,0);

logujy = j(ngroup,1,0); /* (y --- the envelope/proposal function) */

ujstar = j(ngroup*nq,1,0);
c1 = j(ngroup,1,0);
c2 = 4;
etaij = j(N,1,0);
uij   = j(N,1,0);
wij   = j(N,1,0);
logujxp1 = j(N,1,0);
ujcount = j(ngroup,1,0);

/* calculate c1 */

c = 1;
e = 1;
f = 1;
g = nq;

Do i = 1 to ngroup;

/* compute the ordinate of ggam --- the envelope function at the mode */

logujy[c:e] = - ((nq/2) * log (2 * 3.14))
        - ((nq/2) * log (det(c2 * varuj[f:g,1:nq])));
```

```
/* compute the ratio c1 */

c1[c:e] = logujx[c:e] - logujy[c:e];

c = c + 1;
e = e + 1;
f = f + nq;
g = g + nq;

end;

/* generate ujstar from yuj --- the envelope function */

a = 1;
c = 1;
e = 1;
f = 1;
g = nq;
nijth = 0;

Do i = 1 to ngroup;

ratio = 0;

u = 1;
ijth = a - 1 + nj[c:e];        /* size index              */
nijth = ijth - a + 1;

CholeskU[f:g,1:nq] = root(c2*varuj[f:g,1:nq]);

Do while (ratio < u);

nordev = j(nq,1,0);
x = 1;

do k = 1 to nq;
      nordev[x:x] = rannor(seed1);
      x = x + 1;
end;

ujstar[f:g] = uj[f:g] + CholeskU[f:g,1:nq]`*nordev[1:nq];
```

```
/* compute the ordinate of ujstar using guj ---- the envelope function */

logujy[c:e] =   - ((nq/2) * log (2 * 3.14))
            - ((nq/2) * log (det(c2 * varuj[f:g,1:nq])))
            - ((ujstar[f:g]-uj[f:g])`
           *inv(c2*varuj[f:g,1:nq])*(ujstar[f:g]-uj[f:g]))/2;

/* the ordinate of ujstar using fuj ---- the true function */

etaij[a:ijth] = ofstgam[a:ijth]+ zij[a:ijth,1:nq]*ujstar[f:g];
uij[a:ijth] = 1/(1+exp(-etaij[a:ijth]));
logujxp1[a:ijth] = ((yij[a:ijth]#log(uij[a:ijth]))
            + ((1-yij[a:ijth])#log(1-uij[a:ijth])));
logujx[c:e] =   sum(logujxp1[a:ijth])
            - (ujstar[f:g]`*inv(Tau)*ujstar[f:g])/2;

/* compute the ratio */

ujcount[c:e] = ujcount[c:e] + 1;
ratio = exp(logujx[c:e] - c1[c:e] - logujy[c:e]);
u = uniform(seed1);

end;

uj[f:g] = ujstar[f:g];

a = a + nj[c:e];
c = c + 1;
e = e + 1;
f = f + nq;
g = g + nq;

end;

totujcnt = 0;
totujcnt = sum(ujcount);
print totujcnt;

end;
finish;

/******  module gbuj ***************/
```

```
/****** module output ***************/

start output (itcount,firstout,ngroup,nf,nq,niter,nsample,
        outgam,outTau,outuj,gam,Tau,uj);

gamt = j(1,nf,0);          /* gam            */
gamt = gam`;

uniqTau = nq * (nq + 1)/2;      /* unique elements of Tau */
elemTau = j(1,uniqTau,0);

ujt = j(1,ngroup*nq,0);       /* uj             */
ujt = uj`;

if (firstout = itcount) then do;
     if (outgam = 1) then create output.gampost from gamt;
     if (outTau = 1) then create output.Taupost from elemTau;
     if (outuj = 1) then create output.ujpost from ujt;
end;

/* Write out the posterior distribution of the parameters     */

if (outgam = 1) then do;
     setout output.gampost;
     append from gamt;
end;

if (outTau = 1) then do;

     index = 1;
     do i = 1 to nq;
          do j = 1 to nq;
               if (i <= j) then do;
               elemTau[1,index] = Tau[i,j];
               index = index + 1;
               end;
          end;
     end;

     setout output.Taupost;
     append from elemTau;
end;
```

```
if (outuj = 1) then do;
      setout output.ujpost;
      append from ujt;
end;

finish;

/******* module output *********/
```

```
/******* main program **********/

itcount = 1;

run import (yij,lev2var,lev1var,nj);
run initial (yij,lev2var,lev1var,nj,xij,zij,ngroup,N,nf,nq,
        inititer,niter,nsample,
        gam,uj,Tau,seed1,outgam,outuj;outTau);

Tau = 1;
uj = j(ngroup*nq,1,0);

firstout = niter - nsample + 1;

Do while (itcount <= niter);
run glimgam (yij,nj,nf,nq,ngroup,N,xij,zij,gam,uj,vgam,inititer,det,
          loggamx,ofstuj);
run gbgam (yij,nj,nf,nq,ngroup,N,xij,zij,gam,uj,vgam,gengam,seed1,
          loggamx,c2gam,cntgam,accgam,ofstuj,inititer);
run ridgeuj  (yij,xij,zij,nj,N,ngroup,nf,nq,gam,Tau,uj,varuj,inititer,
          logujx,ofstgam);
run gbuj  (yij,xij,zij,nj,N,ngroup,nf,nq,gam,Tau,uj,varuj,genuj,seed1,
          logujx,c2uj,cntuj,accuj,ofstgam,inititer);
run gbTau (uj,Tau,nq,ngroup);
If (itcount >= firstout) then do;
     run output (itcount,firstout,ngroup,nf,nq,niter,
              nsample,outgam,outTau,outuj,gam,Tau,uj);
end;
print itcount;
itcount = itcount + 1;
end;


quit;
```

APPENDIX B

APPENDIX B

Approval Letter from the U.S. Department of Education for Accessing Individually Identifiable Survey Data Base

U.S. DEPARTMENT OF EDUCATION
OFFICE OF EDUCATIONAL RESEARCH AND IMPROVEMENT

NATIONAL CENTER FOR EDUCATION STATISTICS

Yuk Fai Cheong
Department of Counseling
College of Education
Michigan State University
East Lansing, Michigan 48824-1034

SEP 6 1994

Dear Mr. Cheong:

I am pleased to inform you that the Department of Education, Michigan State University has met the requirements for accessing the individually identifiable survey data base entitled: "1992 NAEP Trial State Assessment".

The following information is enclosed for your use:

- A signed copy of the license and 4 cop(ies) of the Affidavit of Non-Disclosure, for yourself and three project staff;
- One copy of the data base you requested on four 9-track computer tapes numbered: W-1565, W-01520, W-01483 and W05297; and
- Disk containing documentation.

Please keep the single copy of the Privacy Act of 1974, GEPA, and the NCES Security Procedures, enclosed with your initial licensing application, with the executed license for reference by you and those project staff who will be accessing the data. Also retain a copy of your approved data Security Plan with the executed license. Violations of any of the licensing provisions by any member of your research project staff could result in cancellation.

These data are on loan to Department of Education, Michigan State University for a period not to exceed 5 years commencing with the date of the NCES Commissioner's signature on the license. You have been assigned license number: 940830148. Reference this number in all future correspondence.

If you have any questions, you may call me at (202) 219-1920.

Sincerely,

Alan W. Moorehead
Data Security Officer

Enclosures

WASHINGTON D.C. 20208—

APPENDIX C

APPENDIX C

Approval Letter from the University Committee on Research Involving Human Subjects

MICHIGAN STATE
UNIVERSITY

November 20, 1996

TO: Stephen W. Raudenbush
College of Education
461 Erickson Hall

RE: IRB#: 94-412
TITLE: CORRELATES OF STATE VARIATION IN THE SOCIAL DISTRIBUTION OF ACHIEVEMENT: A BAYSEAN ANALYSIS: METHODOLOGICAL ALTERNATIVES IN THE ANALYSIS OF DATA FROM NAEP
REVISION REQUESTED: 11/05/96
CATEGORY: 1-E
APPROVAL DATE: 09/18/96

The University Committee on Research Involving Human Subjects' (UCRIHS) review of this project is complete. I am pleased to advise that the rights and welfare of the human subjects appear to be adequately protected and methods to obtain informed consent are appropriate. Therefore, the UCRIHS approved this project and any revisions listed above.

RENEWAL: UCRIHS approval is valid for one calendar year, beginning with the approval date shown above. Investigators planning to continue a project beyond one year must use the green renewal form (enclosed with the original approval letter or when a project is renewed) to seek updated certification. There is a maximum of four such expedited renewals possible. Investigators wishing to continue a project beyond that time need to submit it again for complete review.

REVISIONS: UCRIHS must review any changes in procedures involving human subjects, prior to initiation of the change. If this is done at the time of renewal, please use the green renewal form. To revise an approved protocol at any other time during the year, send your written request to the UCRIHS Chair, requesting revised approval and referencing the project's IRB # and title. Include in your request a description of the change and any revised instruments, consent forms or advertisements that are applicable.

PROBLEMS/ CHANGES: Should either of the following arise during the course of the work, investigators must notify UCRIHS promptly: (1) problems (unexpected side effects, complaints, etc.) involving human subjects or (2) changes in the research environment or new information indicating greater risk to the human subjects than existed when the protocol was previously reviewed and approved.

If we can be of any future help, please do not hesitate to contact us at (517)355-2180 or FAX (517)432-1171.

Sincerely,

David E. Wright, Ph.D.
UCRIHS Chair

DEW:bed

cc: Yuk Fai Cheong

OFFICE OF
RESEARCH
AND
GRADUATE
STUDIES

University Committee on
Research Involving
Human Subjects
(UCRIHS)

Michigan State University
232 Administration Building
East Lansing, Michigan
48824-1046

517/355-2180
FAX: 517/432-1171

REFERENCES

LIST OF REFERENCES

Annie E. Casey Foundation. (1994). Kids count data book: State profiles of child well-being. Washington, DC: Center for Study of Social Policy.

Anderson, D. A., & Aitkin, M. (1985). Variance component models with binary response: Interviewer variability. Journal of the Royal Statistical Society, Series B, 47, 203-210.

Archer, E. (1993). New equations: The Urban Schools Science and Mathematics Program. New York, NY: Academy for Educational Development.

Becker, H. J. (1990). Curriculum and instruction in middle-grade schools. Phi Delta Kappan, 71, 450-457.

Besag, J., Green, P., Higdon, D., & Mengersen, K. (1995). Bayesian computation and stochastic systems. Statistical Science, 10(1), 3-66.

Best, N, Cowles, M. K., & Vines, K. (1995). CODA Convergence Diagnosis and Output Analysis Software for Gibbs sampling output. Cambridge: MRC Biostatistics Unit.

Box, G. E. P., & Tiao, G. C. (1973). Bayesian inference in statistical analysis. New York: Wiley.

Breslow, N. W., & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. Journal of the American Statistical Association, 88, 9-25.

Bruckerhoff, C. (1995). Life in the bricks. Urban Education, 30(3), 317-336.

Bryk, A. S., Raudenbush, S. W., & Congdon, R. T. (1996). Hierarchical linear and nonlinear modeling with the HLM/2L and HLM/3L Programs. Chicago: IL: Scientific Software International, Inc.

Carter, R. L. (1995). The unending struggle for equal educational opportunity. Teachers College Records, 96(4), 619-626.

Casella, G., & George, E. I. (1992). Explaining the Gibbs sampler. The American Statistician, 46(3), 167-174.

Chambers, D. L., (1994). The *right* algebra for all. Educational Leadership, 51(6),

85-86.

Council of Chief State School Officers. (1993). Analysis of state education indicators: State profiles and NAEP results related to state policies and practices. Washington: DC: Author.

Cowles, M. K. (1994). Practical issues in Gibbs sampler implementation with application to Bayesian hierarchical modeling of clinical trial data. Unpublished doctoral dissertation, University of Minnesota.

Dagpunar, J. (1988). Principles of random variate generation. Oxford: Clarendon Press.

DeMaris, A. (1992). Logit modeling: Practical applications. (Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07-086). Newbury Park, CA: Sage.

Dempster, A. P., Laird, N. M., & Rubin D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). Journal of the Royal Statistical Society, Series B, 39, 1-18.

Dellaportas, P., & Smith, A. F. M. (1993). Bayesian inference for generalized linear and proportional hazards via Gibbs sampling, Applied Statistics, 42, 443-459.

Draper, D. (1995). Inference and hierarchical modeling in the social sciences. Journal of Educational and Behavioral Statistics, 20(2), 115-147.

Epstein, J. L., & McPartland, J. M. (1988). Education in the middle grades: A national survey of practices and trends - survey and telephone interview forms. Baltimore: John Hopkins University, Center for Research on Elementary and Middle Schools.

Edwards, T. G. (1992) Current reform efforts in mathematics education. Washington, DC. (ERIC Document Reproduction Service No. ED 372 969).

Fahrmeir, L., & Tutz, G. (1994). Multivariate statistical modeling based on generalized linear models. New York: Springer-Verlag.

Gamoran, A. (1987). The stratification of high school learning opportunities. Sociology of Education, 60, 135-155.

Gelfand, A. E. (1994). Gibbs sampling. To appear in the Encyclopedia of Statistical Sciences.

Gelfand, A. E., & Smith, A. F. M. (1990). Sampling based approaches to calculating marginal densities. Journal of the American Statistical Association, 85, 398-409.

Gelfand, A. E., Hills, S., Racine-Poon, A. & Smith, A. F. M. (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling. Journal of the American Statistical Association, 85, 972-985.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). Bayesian data analysis. New York: Chapman & Hill.

Gemen, S., & Gemen, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 6, 721-741.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In J. M. Bernardo, J. O. Berger, A. P. David, & A. F. M. Smith (Eds.) Bayesian Statistics 4 (pp. 169-193). Oxford: Clarendon Press.

Gibbons, R. D., & Hedeker, D. (1994). Application of random-effects probit regression models. Journal of Consulting and Clinical Psychology, 62, 285-296.

Gilks, W. R. (1996). Full conditional distributions. In W. R. Gilks, S. Richardson, & D. J. Spiegelhalter Eds.) Markov Chain Monte Carlo in practice (pp. 75-88). London: Chapman & Hill.

Gilks, W. R., Best, N. G., & Tan, K. K. C. (1995). Adaptive rejection Metropolis sampling within Gibbs sampling. To appear in Applied Statistics.

Gilks, W. R., & Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. Applied Statistics, 41, 337-348.

Gilks, W. R., Clayton, D. G., Speigelhalter, D. J., Best, N. G., McNel, A.J., Sharples, L. D., Kirby, A. J. (1993). Modeling complexity: Applications of Gibbs sampling in medicine. Journal of the Royal Statistical Society, Series B, 55, 39-52.

Goldstein, H. (1991). Nonlinear multilevel models, with an application to discrete response data. Biometrika, 78(1), 45-51.

Goldstein, H., & Rasbash, J. (1995). Improved approximations for multilevel models with binary response. Journal of the Royal Statistical Society, Series B, 55, 39-52

Gordon, E. W. (1967). Equalizing education opportunity in the public school. IRCB Bulletin, (3)5, 1-3.

Haller, E. J., Monk, D. H., Bear, A. S., Griffith, H, & Moss, P. (1990). School size and program comprehensiveness: Evidence from High School and Beyond. Educational Evaluation and Policy Analysis, 12(2), 109-120.

Hallinan, M. T. (1992). The organization of students for instruction in the middle school. Sociology of Education, 65, 114-127.

Hallinan, M. T. (1996). Educational processes and school reform. In A. C. Kerckhoff (Ed.), Generating social stratification: Toward a new research agenda (pp. 153-169). Boulder, CO: Westview Press.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. Biometrika, 57, 97-109.

Hawley, W. D. (1983). Achieving quality integrated education--with or without federal help. Phi Delta Kappa, 64(5), 334-338.

Hedeker, D. R., & Gibbons, R. D. (1994). A random-effects ordinal regression model for multilevel analysis. Biometrics, 50, 933-944.

Hedeker, D. R., & Gibbons, R. D. (in press). MIXOR: a computer program for mixed-effects ordinal regression analysis. To appear in Computer Methods and Programs in Biomedical Research.

Hedges, L. V., Laine, R. D., & 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.

Heidelberger, P., & Welch, P. (1983). Simulation run length control in the presence of an initial transient. Operation Research, 31, 1109-44.

Hosmer, D. W., & Lemeshow, S. (1989). Applied logistic regression. New York: John Wiley.

Jetter, A. (1993). Mississippi learning. New York Times Magazine, 21, 28-72.

Kamii, M. (1990). Opening the algebra gate: Removing obstacles to success in college preparatory mathematics courses. Journal of Negro Education, 59(3), 392-405.

Karim, M. R. (1991). Generalized linear models with random effects. Unpublished doctoral dissertation. Johns Hopkins University.

Karim, M. R., & Zeger, S. L. (1992). Generalized linear models with random effects:

Salamander mating revisited. Biometrics, 48, 631-644.

Kuk, A. Y. C. (1995). Asymptotically unbiased estimation in generalized linear models with random effects. Journal of Royal Statistical Society, Series B, 57, 395-407.

Lee, V. E. & Smith, J. B. (1995). Effects of high school restructuring and size and early gains in achievement and engagement. Sociology of Education, 68, 241-270.

Lindstrom, M. J., & Bates, D. M. (1990). Nonlinear mixed effects models for repeated measures data. Biometrics, 46, 673-687.

Longford, N. T. (1988). VARCL: Software for variance component analysis of data with hierarchically nested random effects (maximum likelihood). Princeton: Educational Testing Service.

Longford, N. T. (1994). Logistic regression with random coefficients. Computational Statistics & Data Analysis, 17, 1-15.

Longford, N. T. (1995). Random coefficient models. In G. Arminger, C. C. Clogg, M. E. Sobel (Eds.), Handbook of statistical modeling for the social and behavioral sciences (pp. 519-577). New York: Plenum Press.

MacIver, D. J., & Epstein, J. L. (1995). Opportunities to learn: Benefits of algebra content and teaching-for-understanding instruction for 8th-grade public school settings. Baltimore: The Johns Hopkins University, Center for the Social Organization of Schools.

McCullagh, P., & Nelder, J. A. (1989). Generalized Linear Models (2nd ed.). London: Chapman & Hill.

McGilchrist, C.A. (1994). Estimation in generalized mixed models. Journal of Royal Statistical Society, Series B, 56, 61-69.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., & Teller, A. H. (1953). Equations of state calculations by fast computing machines. Journal of Chemical Physics, 21, 1087-1091.

Mohadjer, L. Y., Rust, K. F., Smith, V., & Severynse, J. (1993). Sample Design and Selection. In E. G. Johnson, J. Mazzeo, & D. L. Kline , Technical report of the NAEP 1992 Trial State Assessment Program in mathematics (pp. 35-86). Washington, DC: National Center for Education Statistics.

Monk, D. H. (1994). Incorporating outcome equity standards into extant systems of

educational finance. In R. Berne & L. O. Picus (Eds.), Outcome equity in education (pp. 224-246). Thousand Oaks, CA: Corwin Press.

Monk, D. H., & Haller, E. J. (1993). Predictors of high school academic course offerings: The role of school size. American Educational Research Journal, 30(1), 3-21.

Moses, R. P. (1994). Remarks on the struggle for citizenship and math/science literacy. Journal of Mathematical Behavior, 13, 107-111.

Moses, R. P., Kamii, M., Swap, S. M., & Howard, J. (1989). The Algebra Project: Organizing in the spirit of Ella. Harvard Educational Review, 59, 423-443.

Mullis, I. V. S. (1991). The NAEP guide: A description of the content and methods of the 1990 and 1992 assessments. (Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress, 1990 and 1992).

Mullis, I. V. S., Jenkins, F., & Johnson, E. G. (1994). Effective schools in mathematics. Washington, DC: National Center for Education Statistics.

National Center for Education Statistics. (1994). The Condition of Education 1994. Washington, D.C.: United States Government Printing Office.

National Council of Teachers of Mathematics. (1989). Curriculum and evaluation standards for school mathematics, Reston, VA: NCTM.

National Council of Teachers of Mathematics. (1993). Algebra for the Twenty-First Century. Proceedings of the August 1992 Conference, Reston, VA: NCTM.

Nelder, H. A., & Wedderburn, R. W. M. (1972). Generalized linear models. Journal of the Royal Statistical Society, Series A, 135, 370-84.

Oakes, J. (1989). What educational indicators? The case for assessing the school context. Educational Evaluation and Policy Analysis, 11(2), 181-199.

Oakes, J. (1990). Multiplying inequalities: The effects of race, social class, and tracking on opportunities to learn math and science. Santa Monica, CA: RAND.

O'Day, J. A., & Smith, M. S. (1993). Systemic reform and educational opportunity. In S. Fuhrman (Ed.), Designing coherent education policy: improving the system (pp. 250-312). San Francisco: Jossey-Bass Publishers.

Odell, P. L., & Feiveson, A. H. (1966). A numerical procedure to generate a sample covariance matrix. Journal of the American Statistical Association, 61, 198-203.

Olson, L. (1994, May 18). Algebra focus of trend ro raise stakes. Education Week, pp. 1, 11.

Pelgrum, W. J., Voogt, J., & Plomp, T. (1995). Curriculum indicators in international comparative research. In Measuring the quality of schools: Indicators of education systems (pp. 80-102). Washington, DC: OECD Publications and Information Center.

Porter, A. C. (1991). Creating a system of school process indicators. Educational Evaluation and Policy Analysis, 13(1), 13-30.

Porter, A. C. (1994). National standards and school improvement in the 1990s: Issues and Promise. American Journal of Education, 102, 421-449.

Prosser, R., Rasbash, J. & Goldstein, H. (1991). ML3: software for 3-level analysis: Users Guide. London: Institute of Education.

Raftery, A. L., & Lewis, S. (1992). How many iterations in the Gibbs sampler? In J. M. Bernardo, J. O. Berger, A. P. David, & A. F. M. Smith (Eds.) Bayesian statistics 4 (pp. 763-773). Clarendon Press, Oxford, UK.

Raftery, A. L., & Lewis, S. (1996). Implementing MCMC. In W. R. Gilks, S. Richardson, & D. J. Spiegelhalter (Eds.) Markov Chain Monte Carlo in Practice (pp.115-130). London: Chapman & Hill.

Raudenbush, S. W. (1988). Educational applications of hierarchical linear models: A review. Journal of Educational Statistics, 13, 85-116.

Raudenbush, S. W. (1993). Posterior modal estimation for hierarchical generalized linear models with application to dichotomous and count data. Unpublished manuscript now under review.

Raudenbush, S. W., Cheong, Y. F., & Fotiu, R. P. (1995). Synthesizing cross-national classroom effect data: Alternative models and methods. In M. Binkley, K. Rust, and M. Winglee (Eds.) Methodological issues in comparative international studies: The case of reading literacy (243-286). Washington, D.C.: National Center for Educational Statistics.

Raudenbush, S. W., Fotiu, R. P., Cheong, Y. F., & Ziazi, Z. M. (1996). Inequality of access to educational opportunity: A national report card for eighth-grade math.

Unpublished manuscript now under review.

Raudenbush, S. W., Kasim, R. M., Eamsukkawat, S., & Miyazaki, Y. (1996). Social origins, schooling, adult literacy, and employment: Results from the National Adult Literacy Survey. Unpublished manuscript now under review.

Ripley, B. (1987). Stochastic simulation. New York: John Wiley and sons.

Rodriguez, G., & Goldman, N. (1995). An assessment of estimation procedures for multilevel models with binary responses. Journal of Royal Statistical Society, Series A, 56, 73-89.

SAS Institute, Inc. (1989). SAS/IML: Usage and Reference, Version 6. Cary, NC: SAS Institute, Inc.

Schall, R. (1991). Estimation in generalized linear models with random effects. Biometrika, 40, 719-727.

Seltzer, M. H., Wong, W. H., & Bryk, A. S. (1996). Bayesian analysis in applications of hierarchical models: Issues and methods. Journal of Educational and Behavioral Statistics, 21(2), 131-167.

Searle, S. R., Casella, G., & McCulloch, C. E. (1992). Variance Components. New York: John Wiley & Sons.

Silva, C. M., & Moses, R. P. (1990). The Algebra Project: Making middle school mathematics count. Journal of Negro Education, 59(3), 375-391.

Silver, E. A., (1995). Rethinking "algebra for all". Educational Leadership, 52, 30-33.

Sørenson, A. B., & Hallinan, M. T. (1977). A reconceptualization of school effects. Sociology of Education, 50, 273-289.

Spiegelhalter, D., Thomas, A., Best, N., & Gilks, W. (1994). BUGS Manual 0.30, Cambridge: MRC Biostatistics Unit.

Steen, L. A., (1992). Does everybody need to study algebra? Basic Education, 37(4), 9-13.

Stevenson, D. L., Schiller, K. S., & Schneider, B. (1994). Sequences of opportunities for learning. Sociology of Education, 67, 184-198.

Stiratelli, R., Laird, N., & Ware, J. H. (1984). Random-effects models for serial

observations with binary response. Biometrics, 40, 961-971.

Tanner, M. A. (1993). Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions. 2nd ed. New York: Springer-Verlag.

Tate, W. F. (1994). Mathematics standards and urban educational Is this the road to recovery. The Educational Forum, 58, 380-390.

Tierney, L. (1994). Markov chains for exploring posterior distributions. The Annals of Statistics, 22(4), 1701-1762.

Useem, E. L. (1992). Getting on the fast track in mathematics: School organizational influences on math track assignment. American Journal of Education, 100(3), 325-393.

Usiskin, Z. (1987). Why elementary algebra can, should, and must be an eight-grade course for average students. Mathematics Teacher, 80, 423-437.

Wei, G. C. G., & Tanner, M. A. (1990). A Monte Carlo implementation of EM algorithm and the poor man's data augmentation algorithms. Journal of the American Statistical Association, 85, 699-704.

Wolfinger, R., & O'Connell, M. (1993). Generalized linear mixed models: A pseudo-likelihood approach. Journal of Statistical Computing and Simulation, 48, 233-243.

Yang, M. (1994). A Simulation Study for the Assessment of the Non-linear Hierarchical Model Estimation via Approximate Maximum Likelihood. Unpublished manuscript.

Zeger, S. L., Liang, K. Y., & Albert, P.S. (1988). Models for longitudinal data: A generalized estimating equation approach. Biometrics, 44, 1049-1060.

Zeger, S. L., & Karim, M.R. (1991). Generalized linear models with random effects: a Gibbs sampling approach. Journal of the American Statistical Association, 86, 79-86.