

COMMON AND TEXTBOOK FOIL GROUPINGS: A SOCIAL NETWORK APPROACH TO
DISTRACTOR ANALYSIS

By

Leslie Pearlman

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Measurement and Quantitative Methods

2011

ABSTRACT

COMMON AND TEXTBOOK FOIL GROUPINGS: A SOCIAL NETWORK APPROACH TO
DISTRACTOR ANALYSIS

By
Leslie Pearlman

This dissertation examines the patterns and types of mistakes students make on a large-scale mathematics assessment, and puts these patterns into perspective based on the textbook used and the specific content covered in a student's classroom. In order to investigate the nature of these patterns, I employ a method of analysis originally designed for modeling the social interactions/networks of actors attending different events. Once these patterns are defined, I assign content, process and skill codes to each item, following the methods of the Rule Space Method, and identify different "Knowledge States" for each position of items and students within a school. These Knowledge States are compared within schools using the same textbooks, and between schools using different texts.

The analyses carried out in separate phases, addresses two main hypotheses. The first phase identifies the different Knowledge States for each school in 4 districts, each using a unique textbook and/or curriculum; all aligned to the State of Michigan's Grade-Level-Content-Expectations. Hypothesis I suggests that the positions of items and students in each school are meaningful, and not likely to occur by chance alone. In order to examine Hypothesis I, I examine the statistical significance of the CliqueFinder algorithm's results, and then provide an analysis of the distractors chosen by all students within this sample, and compare these to the results for each school and textbook combination.

The second phase addresses Hypothesis II, which the Knowledge States are consistent across schools using the same textbooks, and different between schools using different texts. I provide descriptive analyses for each school related to the reported content coverage and textbook use, specifically the use of textbooks specifically designed to align with National and individual state standards.

Copyright by
LESLIE PEARLMAN
2011

Dedicated to my parents, for their endless love, support and encouragement, through everything.

ACKNOWLEDGMENTS

In my third year of graduate studies, I took a social network class with Dr. Ken Frank, who is now my advisor, dissertation supervisor, and most importantly my mentor. That class altered the direction of my studies forever. I had been searching for a connection to my previous studies in economics, and Ken's social networks class illuminated the connections between instrument design and modeling individual behavior. Ken provided me with numerous opportunities to work with so many different projects and data, and enabled me to think creatively and rigorously about examining the relationships between individuals and behavior. I'm so thankful to Ken for introducing me to social network analysis, but mostly for his endless support and encouragement, and most importantly for helping me to learn to *think* about creative and novel solutions to existing problems.

One of the most important opportunities, Ken provided me, was the conversation he had with Dr. Joseph Martineau, which eventually lead to this dissertation. It was their conversation, and Joseph's willingness to allow me to use the MEAP data to perform these analyses, and for that, there are no words to describe how grateful I am. Thank you both.

This dissertation reflects the multiple years of experience with test design, curriculum analysis and social network models. I am so thankful for Karen Glickman, who introduced me to Jackie Babcock, and the TIMSS staff, which lead to my first assistantship in the College of Education. During the first five years of my graduate studies, my work with Dr. William Schmidt and Dr. Richard Houang provided me with first-hand experience in building, writing and analyzing large-scale mathematics and science assessments and

examining these results in the context of the curriculum and content students see in the classroom. Bill and Richard were my first mentors in my graduate studies, and I'm forever indebted to them for their endless support and for everything they taught me. It is also because of Bill and Richard that I was able to complete these analyses, and I'm so thankful they shared the PROM/SE teacher data, enabling the main results of this dissertation. I also want to thank Dr. Lee Cogan and Dr. Neelam Kher, Andrew Middlestead, Jimmy Zhou and Kathy Wight, for their endless support and for all they have taught me.

I am so grateful for the support (and understanding) of my committee, Dr. Mark Reckase, and Dr. Joseph Martineau, and am so thankful for their time and contributions to this work. Dr. Kimberly Maier, my original academic advisor has always been there for me, to listen and to provide support, and I appreciate all of the academic and personal support she has given me since 2004.

I mentioned Karen Glickman and Jackie Babcock, above and want to thank them specifically for all of the support, emotionally and otherwise over these years. I decided to come to MSU because of Karen and her husband, Ken, and I decided to stay because of people like Jackie and her husband Phil.

No dissertation comes without stress and emotional mishaps, and I would not have made it this far without the love and support of my family and friends. Vickie and Harvey Hohauser and Irv and Glorianne Pearlman, you are the best parents anyone could hope for, thanks for everything. I'd also like to acknowledge the support of my siblings, Eric, Todd and Jay Hohauser, the best brothers a girl could have, and to Melissa and Liz, the best sister-in-laws a girl who loves her brothers could have. Thanks too, to my grandparents

Beverly Angellotti and Edward and Mary Angellotti, and Selma Silverman; my Aunts, Rita, Marjorie Suzanne and Gail, and my late Uncles Edward and Eugene. I love you.

I also want to mention some of my friends who believed in me when I didn't, and who supported me when I needed it. I need to thank Michelle Robbins for her love and support and for taking the endless hours to edit and re-edit this work. I couldn't have done this without her! I want to thank Kate Roach and Danielle Roberty for the ongoing phone support. Kate listened to me every day for the past several years, and I don't know what I'd do without her. Danielle survived the tree incident, with me and so much more. I want to also thank my best friend Jay Scofield, he inspires me every day with his spirit. I want to also thank (my pseudo-big brother) Dustin Sprigg, he has been an amazing source of support. I'd like to thank my good friend, Derek Moore for encouraging me to keep going, and when I couldn't write any more, to ride my bicycles. I want to thank my friends Renatas Tukys, Matas Anuzis, Al Kasputis, Aras Butkunas, Peter Cohen and Mike Burke for providing comic relief and emotional support. Last but not least, The Peanut Barrel and their amazing staff who I consider to be some of my very favorite friends and biggest supporters.

I'd like to thank my friend and colleague, Venessa Keesler, who provided so much support and always a listening ear, especially during comprehensive exams and while writing this dissertation. Thanks Venessa.

Finally, I'd like to acknowledge every teacher, professor and dean who gave me a hard time, who believed in me when I didn't, and who gave me the foundation to do whatever I set my mind to. Thanks especially to Mr. Tim Domanski, Dr. James Stephens, and Dr. Lee Coppock.

TABLE OF CONTENTS

LIST OF TABLES	XI
LIST OF FIGURES.....	XII
CHAPTER 1 : BACKGROUND	1
Curriculum and Assessment Reform.....	2
Current methods of distractor analysis.....	4
Comprehensive Differential Item Functioning	5
Item Response Theory Model.....	6
Differential Distractor Functioning.....	7
The Rule-Space Method	8
New Directions: Social Network Models.....	10
CHAPTER 2 : DATA, METHODS AND MEASURES.....	14
Data	14
Michigan Educational Assessment Program (MEAP) Data	14
Promoting Rigorous Outcomes in Math and Science Education (PROM/SE) Data	17
Textbook Research and Background	22
Methods:.....	27
Phases I and II – Social Network Models and Correspondence Analysis	27
Phase III –Teacher Consultations, Factor Analysis, and BCLM Logistic Regression.....	34
Principal Factor Analysis and Logistic Regression.....	36
Measures	38
Social Network Analysis	38
Principal Factor Analysis and Logistic Regression.....	40
CHAPTER 3 : ANALYSIS.....	43
Phase I – Common Foil Groups.....	43
Phase II – Textbook Foil Groups and Correspondence Analysis	53
Phase III – Principal Factor Analysis and BCLM Logistic Regression.....	55
Factor Analysis of Process and Skill Codes	56

BCLM Logistic Regression.....	57
CHAPTER 4 : RESULTS.....	60
Phase I – Common Foil Groups.....	60
Phase II – Textbook Foil Groups	62
Correspondence Analysis.....	62
Phase III – Principal Factor Analysis and BCLM Logistic Regression.....	70
Teacher Consultations	70
Textbook and CFG expectations	70
Factor Analysis	74
BCLM Logistic Regression.....	78
Everyday Math.....	84
Investigations	84
Harcourt Math	85
Trailblazers.....	85
CHAPTER 5 : DISCUSSION	86
Conclusions and Discussions.....	86
Limitations.....	89
Future Directions	90
REFERENCES	91

LIST OF TABLES

Table 2.1: Grade Level Content Expectations for the State of Michigan	16
Table 2.2 : Statewide Demographics	19
Table 2.3 : Sample Demographics – Student Demographics by District and Textbook.....	20
Table 2.4 : Odds Ratio Table for Position Membership and Event Participation.....	31
Table 2.6 : Item and Distractor Attribute Codes	41
Table 3.1: Odds Ratio and Likelihood Ratio test for Common Foil Groups	44
Table 3.2: Item and Distractor Information by Common Foil Group.....	45
Table 3.3: Odds Ratio and Likelihood Ratio test for Common Foil Groups	54
Table 4.1: Two way Contingency Table for CFG by Textbook	64
Table 4.2: Eigenvalues and Inertial for all Dimensions.....	64
Table 4.3: Correspondence Analysis Results.....	65
Table 4.4: Eigenvalues of the Reduced Correlation Matrix: Total=6.06 Average=0.55	75
Table 4.5: Rotated Factor Structure.....	78
Table 4.6: Model (1) Fit Statistics	79
Table 4.7: Model (1) Null Hypothesis, Beta=0	79
Table 4.8: Odds Ratios for CFG Membership by Textbook with Wald Confidence Intervals.....	81
Table 4.9: Model (2) Fit Statistics	82
Table 4.10: Model (2) Null Hypothesis, Beta=0	82
Table 4.11: Odds Ratios for Factor Association by Textbook with Wald Confidence Intervals.....	83

LIST OF FIGURES

Figure 2.1: Example of One-Mode Data	39
Figure 2.2: Example of Two-Mode Data	39
Figure 4.1: Correspondence Analysis of CFG by Textbook – Dimension 1 vs. Dimension 2	69
Figure 4.2: Scree and Variance Explained Plots.....	76

CHAPTER 1 : BACKGROUND

Over the past twenty years in the United States, the use of and dependency on standardized testing has, in some ways, outgrown our ability to fully understand and use the data generated from such assessments. Requirements of these data go above and beyond providing parents, teachers, and schools with information on how the student performs on a particular math assessment. These assessments must additionally compare students, classrooms, schools, and districts on the same metric, measure the efficacy of the State's curriculum, and provide a snapshot of performance for use in the longitudinal studies of progress. Decisions regarding funding and staffing are also tied to these scores and their interpretations.

A large part of making sure assessments contain and measure the proper content and set of skills, specifically with multiple-choice items, involves the writing of items and distractors (incorrect choices). Distractors should represent incorrect answers that less proficient students or students with other, perhaps latent, learning anomalies tend to choose. It follows that, in addition to mathematics content, the item writers should understand the common computational mistakes the target test takers are most likely to make.

Current analysis does not take into account the information contained in the incorrect answers students choose. Scores are reported in terms of items answered either correctly or not but, by dichotomizing the item into right or wrong, key information becomes cloudy or lost. When considered along with student profiles, wrong answers and the patterns of distractors chosen should provide additional information to further inform teaching and curriculum practices.

The patterns of distractor choices on a mathematics assessment should be analyzed in context of existing literature on how students learn and process mathematics knowledge, as well as the curriculum and content coverage to which the students are exposed. Research on content, pedagogy, and cognitive processes can provide rich insight into the patterns of wrong answers. This research, when combined with the expertise of teachers in the field, can yield powerful tools to help us understand both the nature of students' mathematical learning and that of common mistakes and cognitive correlates among content, processing, or skill categories.

One outcome of such research efforts was the formation of national standards for mathematics instruction and assessment. This led to the production of textbooks and curriculum specifically designed to align with these standards, which are easily adapted to individual state standards.

The following section details several reform efforts and the development of the textbooks in use for many students in the United States. Next, I review the current methods of distractor analysis, and suggest an alternative method of analysis. Finally, I outline the research hypothesis examined in this dissertation.

Curriculum and Assessment Reform

Recognizing the need for reform related both to curriculum and assessment, several organizations including the National Council of Teachers of Mathematics (NCTM), the National Council of Supervisors of Mathematics (NCSM), the Mathematical Sciences Education Board (MSEB), and the National Research Council (NRC), sought to create standards for mathematics curriculum, pedagogy and assessment (NRC, Shavelson, & Towne, 2002).

These studies produced standards for mathematical content, outlining specific tasks and content coverage across grade-levels coordinated with standards for aligning assessments to the curriculum; they paid specific attention to the tasks and content contained in the curriculum standards. Publications emerging from these studies include: “The Curriculum and Evaluation Standards”(NCTM 1989), “Everybody Counts – A report to the Nation on the Future of Mathematics Education”(NRC and MSEB 1989), “Reshaping School Mathematics – A Philosophy and Framework for Curriculum”(MSEB 1990), and “Professional Standards for Teaching Mathematics”(NCTM 1991).

Additionally, these studies provided the impetus for the NRC’s funding of three institutions’ production of textbooks and curriculum specifically designed to reflect the NCTM Standards. The University of Chicago produced *Everyday Mathematics*, The University of Illinois, Chicago designed *Math: Trailblazers*, and TERC in Cambridge, MA created *Investigations in Number, Data and Space*.

These three curricula share a common goal of aligning the content of the textbooks to the standards and presenting rigorous mathematical content in broader and more practical applications and contexts. Specifically, curricula should: connect students’ experiences to mathematics; challenge and engage students of varying mathematical prowess; help students explore alternative and varying problem-solving strategies; vary instruction based on different student abilities; assist teachers in presenting differing mathematics concepts to students; and provide various assessment instruments and tools. Although development of these curricula began in the 1980’s, the evolution of the textbooks occurred one grade-level at a time. Through revisions and field testing, the final versions emerged, and represent the editions currently in use (COMAP 2003).

Various studies show the positive impact of standards-based reform, especially when combined with standards-based curricula (Carpenter, Fennema et al. 1989; Cobb, Wood et al. 1991; Hiebert and Wearne 1993; Fennema, Carpenter et al. 1996; Hiebert, Carpenter et al. 1996; Hiebert and Wearne 1996; Fuson and Briars 2000). The reform textbooks discussed in this dissertation align with the National Standards and contain innovative approaches for introducing content and enhancing problem-solving skills. Most State standards are derived from those, and assessments are aligned to the State standards. With the new ways of teaching mathematics, and thus new methods of student problem-solving, the question regarding assessment becomes: do the wrong items represent all of the types of mistakes students can make with respect to known misconceptions and new curricula?

New curricula designed to align with State-level content specifications and benchmarks may additionally complicate the issue of distractor composition by test writers. As an example, students using a curriculum that teaches them to add and subtract using methods other than carrying and borrowing will make different computational mistakes based on the way the content was presented in the classroom than fellow students using a different curriculum. This simple example helps illustrate the need to understand how students are using various curricula in order to derive valid distractors for all students, regardless of the curriculum to which they are exposed.

Current methods of distractor analysis

Currently four main methods of distractor analysis exist and each of these methods provides evidence suggesting the importance of examining distractor choices. The

methods differ greatly from one to the next regarding the nature and limitations of the specific analysis.

Comprehensive Differential Item Functioning

The first of the four methods extends the standardization approach (Dorans and Holland 1993) for assessing differential item functioning (DIF) to include the distractor options and omitted responses. This extended approach described in Dorans Schmitt and Bleistein(1992) is called the standardization approach for comprehensive differential item functioning (Cdif). The Dorans and Holland chapter explains a two-part method for identifying and describing DIF with respect to items on an instrument.

We say that an item exhibits DIF if, for two comparable subgroups of a sample (matched on total test score and categorized by some additional characteristic such as gender, ethnicity, educational level, etc.), a different proportion of examinees from the focal group answers correctly compared to the reference group. In order to detect DIF, the authors describe the Mantel-Haenszel (MH) approach. Identifying DIF using this method involves a contingency table based on the proportion of right and wrong answers chosen by both of the groups and testing a null hypothesis of no DIF by comparing the ratios of correct answers for the groups.

The MH approach requires a matching variable for the focal and reference groups. Most often, analysts use a total test score to match students. If we were looking to see if an item functioned differently for females, we would break the sample into male and female groups and then match the students on total test scores. The test statistic for the null hypothesis of no DIF is a ratio of correct/incorrect answers for the reference group to the focal group. If this ratio is equal to one, we do not reject the null.

In order to describe any DIF that may occur, the authors suggest using the standardization approach. Here, empirical item-test regressions are generated and compared for the reference and focal groups. Deviations from the expected performance on an item for students of equal abilities indicate DIF. The standardization method uses all of the information in the item, including missing or omitted responses. Further, the sizes of the groups at different ability levels, i.e. different test scores, need not be equal. The Cdif approach is used as an adjunct to the MH DIF detection in operational testing programs. In the article, the authors use this approach with SAT data to examine the differential speediness of examinees.

Item Response Theory Model

In addition to the standardized Cdif method there are two models, a log-linear model and and Item Response Theory (IRT) based model. Lei et al. provide a brief literature review specifically citing Thissen and Steinberg's Multiple-Choice model (Thissen, Steinberg et al. 1989) and present this model as a general representation of the parametric models currently in use. This model is a modified version of the 2-parameter logistic model used in IRT. The details of this model receive treatment below. These authors also describe some non-parametric models currently used; these models provide similar results to the parametric models, except at the lower ends of the distributions of scores.

The motivation for Thissen and Steinberg in their 1984 article for deriving the Multiple Choice Model was the problem of non-monotonic behavior in the distractors' Item Characteristic Curves, ICC's, when using Bock's (1972) nominal model for an item response. Their final model gives the probability of an examinee with a given ability level choosing a

given option of a given item. The insights from their analysis indicate the importance of close examination of distractor responses.

This method has several serious limitations. The parameter estimation for these models is difficult and computationally burdensome. The authors ran their analysis on only four items. Further, the authors caution against the possibility of still finding non-monotonic curves for examinees based on the rest of that examinee's response pattern. Nonetheless, the authors show that, indeed, the distractors are part of the item.

Differential Distractor Functioning

In the same issue of *Journal of Educational Measurement*, an article presenting a combination of distractor and Differential Item Functioning (DIF) takes a step away from the limiting assumptions of Thissen and Steinberg's model and adds the component of background information of the examinee (Green, Crone et al. 1989). Green et al. argue that perhaps we should look beyond the two-way interaction of ability and answer-choice and include the interaction of background and answer-choice. If, in addition to differences in the distractors chosen by ability, we see group differences in these choices, we should investigate the evidence of some groups being attracted to certain distractors and the possible reasons behind it. They refer to their analysis as the study of Differential Distractor Functioning (DDF). Their model involves separating examinees into distinct ability levels based on the total score of the test. They argue that, with large numbers of items, any dependence from using the total score should be negligible.

The authors use a log-linear model to look at contingency tables of frequencies for option chosen, ability level, and ethnicity. The main effects of their model represent differences in proportions for that subgroup. The interaction effect between ability and

option choice should interest test authors, as it indicates which distractors are attractive to test-takers of different abilities. The significance of the interaction between subgroups and the option choice indicates that a member of a subgroup chooses a certain distractor more than would be expected given his or her ability level.

The hypothesis of interest is that the interaction between subgroup and option choice is not necessary to explain the observed item responses. In order to make a claim for the existence of DFF, it must be demonstrated that data are not explained in a model excluding this interaction and are explained when this term is included in the model.

The three models presented above come with limitations in application and in the conclusions we may draw from them. The Cdif method works well but requires pair wise analysis and a forced matching variable. The log-linear model provides interesting information to the test authors about the value of a given distractor with respect to an ability level. The results from this analysis are inferential. The grouping based upon total-score groups may not be optimal and not all of the item data are used. A large benefit of this model lies in its ability to simultaneously analyze all students and items. The IRT-based model works best with small numbers of items and when the IRT model fits the data. It may be the case, however, that the IRT model fits the keyed response and not the distractor data. This model also hinges on the proper identification of a DIF-free matching variable.

The Rule-Space Method

Tatsuoka et al. developed a method of looking at the processing sub-skills, or attributes necessary to answer items correctly, on an international mathematics assessment (2004). In their study, the authors looked at responses from 20 out of 38

countries that participated in the Third International Math and Science Study-Revisited (TIMSS-R) in 1999. The goal of their study was to illuminate the cognitive processing skills as well as reading or other skills necessary to correctly answer mathematics items. Based on patterns of skills identified by their methodology, the authors identified “knowledge-states” to describe the nature of the responses on the assessment beyond the total number correct.

For their study, the authors convened a panel of experts to look at the items and assign sets of content and processing sub-skill codes to each item. The content codes identify the mathematics content in question. Process attributes identify the processes necessary to correctly answer the item (i.e. Applying Rules in Algebra.) Finally, the skill or item type attribute codes identify the type of item (i.e. Unit Conversion or Estimation.) After each item is coded with all of the appropriate attribute codes, a matrix of all possible patterns of correct answers is created, called a “Q-Matrix”. Each possible set of binary indicators of attributes represents what the authors refer to as “Knowledge-States”. These Knowledge-States represent mastery or non-mastery of the attributes.

The Q-matrix serves as a translation from the latent knowledge-state to the observable pattern of responses for a given student. The rule for the translation is called the rule-space method (RSM) and, given a response pattern, the RSM determines the closest “Knowledge-State” (KS) for a respondent, as well as the probability that the response pattern came from that knowledge-state.

The results from this study show that differences in achievement exist between countries; these differences go beyond content and results are based on process skills that were otherwise undetected using traditional statistical analysis. Differences between

countries with respect to process skills were found to be statistically significant and, upon comparison with other data collected as part of the TIMSS-R, these differences were confirmed by accounts of curriculum coverage and teacher background data.

An additional benefit of the RSM is that multiple assessments may be compared simultaneously by looking at the content, performance, and item types alone. Enabling comparisons of multiple instruments can help to provide evidence to support claims of criterion, content, and construct validity, without having to use or in addition to using traditional equating methods. The results of this research are very promising for test designers, curriculum specialists, and teachers alike. The RSM does not, however, take into account the distractors in multiple choice items. The student answer vectors used to compare with the Q-matrix are binary response vectors and do not take into account the type of mistake made. While the incorrect responses will indicate the items and process types that are problematic for the students, there is no account of the actual mistake made. If this were accounted for, it may provide additional information related to the students' thought processes.

Not only are there limitations to the conclusions we may draw from the four existing methods, but none of these models account for the skills or processes students use to solve the items. Some items may require the use of manipulatives to solve the problem, and some may appear as word problems. The skills and process for these items differ and it is the deficiency in these skills that leads the student to the incorrect answer.

New Directions: Social Network Models

In this dissertation, I use statistical models originally designed for social network studies to model the common behaviors of actors in a given setting. Rather than applying

these models to individuals in an organization, I apply these models to assessment data in order to analyze the responses of 4th graders on a statewide mathematics assessment. It is possible that mistakes on items of different content, i.e. “Number Sense” and “Geometry”, may have something in common when the mistakes on each type of problem are compared. This dependency between mistakes, then, requires analysis capable of taking statistical dependencies into account. Social network models are designed to illuminate underlying social structures, which perhaps account for more variance in behaviors or beliefs than formal structures would suggest. For example, the sharing of resources between elementary school teachers may center more on the scheduled lunch periods than the formal grade-level assignments. Analogous for our assessment example, a student missing two “Number Sense” items and four “Geometry” items may have an underlying issue related to the process or skills necessary to solve the problem rather than the specific content.

This analysis includes three separate and distinct phases. Phase I examines the common distractors chosen across the entire sample of students in the study. This “one-mode” (distractor-to-distractor) analysis identifies the co-occurrences of distractor choices commonly chosen across districts, schools, and textbooks. The clusters of distractors commonly chosen together are referred to as the “Common Foil Groups” (CFGs), and represent the gaps in knowledge (content, skill, and problem-solving processes) observed at the aggregate level.

Phase II explores the relationships between specific textbooks and curricula to which students are exposed and the distractors they chose. A “two-mode” (student-to-distractor) analysis identifies the co-occurrence of students selecting distractors specific to

a textbook or curriculum. The clusters of students and distractors are referred to as the “Textbook Foil Groups” (TFGs), and represent the gaps in knowledge observed at a school-level and unique to specific textbooks.

Following the identification of the TFGs, in Phase III a set of codes is associated with each item and distractor. These codes link the items to specific grade-level content benchmarks for the State. These codes also identify the skills and processes necessary to correctly answer an item. To understand how these item attributes relate to one another and to the CFGs, I use a Principal Factor Analysis to identify the salient factors related to the CFG and TFG formation.

Using the set of ten process and skill codes, I use a Baseline Category Logit Model to estimate the effect that the factors have on predicting CFG membership. I include textbook in a separate model, and this allows us to see the relationships between CFGs and TFGs as well as the differences in which process or skill codes have the largest effects on these group memberships.

These three phases of analysis permit me to examine the following hypotheses:

Hypothesis I: CFGs and TFGs do not occur randomly and are meaningful with respect to known cognitive correlates of mathematical mistake making.

Hypothesis II: TFGs should be consistent among schools using the same textbook and different for schools using other textbooks, while maintaining an overall relationship to the CFGs.

In order to investigate Hypothesis I, I use one- and two-mode network analysis to determine the CFGs and the TFGs. Once these two sets of groups are identified, I examine likelihood statistics to determine whether or not these groups are likely to occur by chance

alone. I also examine the nature of the groups by examining the popularity of each group as well as examining the classical item statistics for distractors making up each CFG and TFG.

Following the Network Analysis, I use logistic regression to examine the effect of item characteristics and textbooks on CFG classifications. In order to model the effects on students' mistakes and item characteristics, I first model CFG membership based on the counts of distractors and the representative factor codes identified with each CFG. In order to model the effects of textbooks on students' mistakes and item characteristics, I model CFG membership based on the same item counts, but do this controlling for textbook. This second model explains the relationship between the CFGs and the TFGs, and therefore textbooks.

The Network Analysis and the Logistic analysis provide statistical evidence to evaluate my set of Hypotheses, and I provide a visual representation of how the CFGs are related to Textbooks using Correspondence Analysis. This analysis allows a representation of how "close" or "distant" CFGs are to one another, and how they are spatially related to textbooks.

CHAPTER 2 : DATA, METHODS AND MEASURES

This chapter describes the various data sets analyzed for this analysis, the methods used, and the variables of interest.

Data

The analysis for this dissertation combines data from Michigan Educational Assessment Program, PROM/SE teacher background data, and data collected from interviews and consultations with several elementary school teachers in three states.

Michigan Educational Assessment Program (MEAP) Data

Data for conducting this analysis should come from a large-scale assessment, based on standards and aligned to the curriculum. A large-scale test given to an entire state will allow the analysis to show how cognitive patterns of mistakes vary between schools and districts, as well as across student demographics. For this specific analysis, it is important not only that the items are well written, but also that they represent what the students have been exposed to in the classrooms (aligned to the curriculum); this ensures that patterns of mistakes are related to content and cognitive processing, rather than unfamiliarity with the content.

In 1969, Michigan's State Board of Education began the statewide assessment program called the Michigan Educational Assessment Program (MEAP). The stated purpose of the program is to "provide information on the status and progress of Michigan education in specified content areas to students, parents, teachers, and other Michigan citizens so individual students are helped to achieve the skills they lack and educators can use the results to review and improve the schools' instructional program" (MEAP 2005 Technical Report). In order to comply with the No Child Left Behind Act (NCLB), the fall of

2005 marked the first time in Michigan when all children in grades three through eight were assessed in mathematics and reading.

According to the State, the proper use of MEAP assessment results can:

- measure academic achievement as compared with expectations, and whether it is improving over time;
- determine whether improvement programs and policies are having the desired effect;
- target academic help where it is needed (MEAP 2005 Technical Report).

In 2005, the testing cycle moved to the fall of each year. This move enabled the state to assess the students based on their prior year of instruction. That is, fourth graders in the fall of 2008 took the version of the MEAP that was based on the State of Michigan's Grade Level Content Expectations (GLCEs) for the third grade curricula. Additionally, beginning in 2005, both the English Language Arts and Mathematics assessments were designed specifically to align with the State's GLCEs, which were approved in 1995 by the Michigan State Board of Education.

The GLCEs are categorized as: (1) Core – Content that is most commonly taught at the grade level; (2) Extended Core – Content commonly taught at grade-level but narrower in scope and/or supportive to core; (3) Future Core – Content expectations previously taught at a higher grade level; will become core-content in 2009-10; and (4) Not Assessed at the State Level (NASL) – GLCEs that are part of the State curriculum but not assessed on the MEAP.

The 2008 MEAP consists of 10 forms and one make-up form. Each form includes a set of two multiple-choice items per Core GLCE. Additionally, each includes a matrix made up of items that are Extended Core (one item per GLCE with a maximum of one-fourth of the extended core GLCEs per form), Future Core (one item per GLCE with a maximum of one future-core GLCE per form) and Field test (as needed).

The MEAP is designed to align with the curriculum to which students are exposed. Curriculum standards and benchmarks outlining the content and knowledge expectations for the content are defined and coded. The design of the MEAP takes these standards and benchmarks into account and aims to assess items specifically matching the content the students see in classrooms. When the test or assessment measures the actual content prescribed by the Standards and Benchmarks, the test is well aligned. Based on version 12.05 of the State's third-grade mathematics GLCEs, four (of five) assessed strands and sub-strands are presented in Table 2.1.

Table 2.1: Grade Level Content Expectations for the State of Michigan

Strand 1 Number & Operations	Strand 3 Measurement	Strand 4 Geometry	Strand 5 Data & Probability
Meaning, notation, place value, and comparisons (ME)	Units and systems of measurement (UN)	Geometric shape, properties, and mathematical arguments (GS)	Data representation (RE)
Number relationships and meaning of operations (MR)	Techniques and formulas for measurement (TE)	Spatial reasoning and geometric modeling (SR)	
Fluency with operations and estimation (FL)	Problem solving involving measurement (PS)		

This analysis uses the response vectors for the fourth graders in 2008 on the mathematics portion of the MEAP. Individual student records include the school code, district code, and student gender, in addition to the response vectors. School and district level demographic and performance data, compiled from the State's website, are also included.

To fully understand the relationships between students and items, specifically how students are most likely to make mistakes, I include teacher data collected by the PROM/SE study. These data show what content teachers covered and to what extent, as well as the textbook and/or curricula used in the school or district. To this end, I limit my inclusion of the MEAP data to those students in participating PROM/SE districts.

Promoting Rigorous Outcomes in Math and Science Education (PROM/SE) Data

Initiated in 2003, the PROM/SE project is funded through a \$35 million partnership agreement from the National Science Foundation (NSF). This project is part of the NSF's Math and Science Partnership (MSP).

Per the mission statement (PROM/SE, 2003):

PROM/SE is a comprehensive research and development effort to improve mathematics and science teaching and learning in grades K-12, based on assessment of students and teachers, improvement of standards and frameworks, and capacity building with teachers and administrators.

Partnering with about 60 school districts, the PROM/SE project collected student data from approximately 300,000 K-12 students and 5,000 teachers from Michigan and Ohio. Data from the teacher questionnaires include background demographics (age,

gender, ethnicity, college attended, college major, etc.) and information on classroom practices, efficacy beliefs, and curriculum/textbooks used.

This analysis includes the data collected from PROM/SE teachers specifically related to the textbook teachers reported using and content coverage reported by the number of class periods devoted to each topic area. These data are reported for the school district level. Pseudonyms were assigned to the districts and schools in order to protect the identity of specific students, teachers, schools, and districts.

Since the MEAP data come from the fall of 2008, I included the teacher data collected in 2006 and 2007 for teachers in Michigan. To investigate the relationship between the groups of students, the items, and the textbooks used, I limited the sample of districts to those reporting 100% implementation of a given textbook for the two years preceding the date of the assessment along with two comparison districts reporting use of multiple textbooks.

Merging the MEAP data with the PROM/SE data – Final Sample

I identified four primary textbooks/curricula that PROM/SE teachers reported using for at least two years prior to the 2008 assessment: (1) Everyday Mathematics (EM), (2) Investigations in Number, Data and Space (INV), (3) Math: Trailblazers (TB), and (4) Math Fundamentals, published by Harcourt Brace Javonovich (HM). I also identified two districts reporting the use of several different mathematics textbooks. These schools serve as the controls for evaluating the performance of the clustering algorithm as well as the conclusions drawn from the results. Table 2.2 provides the student demographics for the state and Table 2.3 contains each of the districts, and by textbook.

Each of the school districts in this analysis varies considerably by size and student demographics both compared to one another and within the textbook categories. These differences suit this analysis because, if the differences in how students choose distractors really do identify cognitive patterns specific to known cognitive associations and/or cognitive patterns specific to curriculum, the results of the clustering algorithm should produce similar patterns regardless of school size and student demographics. Additionally, these variations help to establish how well the algorithm performs for schools of varying size.

Table 2.2 : Statewide Demographics

State of Michigan	
Demographic Subgroup	
Total Students	114239
Male	50.8%
Female	49.2%
Black, not of Hispanic Origin	19.6%
American Indian or Alaskan Native	0.9%
Asian or Pacific Islander	2.8%
Hispanic	5.2%
White, Not of Hispanic Origin	70.2%
Multiracial	1.3%
Students With Disabilities	12.1%
LEP	4.2%
ED	44.1%

Table 2.3 : Sample Demographics – Student Demographics by District and Textbook

Textbook Name	Everyday Math			
District ID	EM1	EM2	EM3	EM4
Total Students	69	55	106	247
Male	52.2%	47.3%	46.2%	49.8%
Female	47.8%	52.7%	53.8%	50.2%
Black, not of Hispanic Origin	56.5%			17.8%
American Indian or Alaskan Native				2.4%
Asian or Pacific Islander	1.4%			12.1%
Hispanic	4.3%		3.8%	6.5%
White, Not of Hispanic Origin	30.4%	100%	96.2%	59.1%
Multiracial	7.2%			2.0%
Students With Disabilities	18.8%	12.7%	17.0%	8.1%
LEP				5.7%
ED	81.2%	41.8%	41.5%	30.4%

Textbook Name	Investigations		
District ID	INV1	INV2	INV3
Total Students	142	424	135
Male	51.4%	46.7%	48.9%
Female	48.6%	53.3%	51.1%
Black, not of Hispanic Origin	0.7%	10.4%	0.7%
American Indian or Alaskan Native		0.7%	0.7%
Asian or Pacific Islander	0.7%	3.5%	0.7%
Hispanic	2.1%	4.0%	5.2%
White, Not of Hispanic Origin	94.4%	79.5%	92.6%
Multiracial	2.1%	1.9%	
Students With Disabilities	14.1%	14.4%	9.6%
LEP		1.9%	
ED	32.4%	28.5%	18.5%

Table 2.3 (cont'd)

Textbook Name	Trailblazers		
District ID	TB1	TB2	TB3
Total Students	214	142	753
Male	58.9%	43.7%	50.1%
Female	41.1%	56.3%	49.9%
Black, not of Hispanic Origin	27.1%		10.2%
American Indian or Alaskan Native		1.4%	2.0%
Asian or Pacific Islander	5.1%	0.7%	0.8%
Hispanic	15.0%	1.4%	3.3%
White, Not of Hispanic Origin	45.8%	96.5%	83.5%
Multiracial	7.0%		0.1%
Students With Disabilities	16.8%	15.5%	12.1%
LEP	2.3%		0.3%
ED	38.3%	36.6%	51.3%
Textbook Name	HM		
District ID	HM1	HM1	
Total Students	76	76	
Male	52.6%	52.6%	
Female	47.4%	47.4%	
Black, not of Hispanic Origin	2.6%	2.6%	
American Indian or Alaskan Native			
Asian or Pacific Islander			
Hispanic	1.3%	1.3%	
White, Not of Hispanic Origin	94.7%	94.7%	
Multiracial			
Students With Disabilities	10.5%	10.5%	
LEP			
ED	53.9%	53.9%	

Table 2.3 (cont'd)				
Textbook Name	Mixed Textbook Use			
District ID	MIX1	MIX2	MIX3	MIX4
Total Students	40	172	128	1004
Male	47.5%	41.9%	43.8%	53.0%
Female	52.5%	58.1%	56.3%	47.0%
Black, not of Hispanic Origin		0.6%	2.3%	44.0%
American Indian or Alaskan Native				1.5%
Asian or Pacific Islander		1.7%	0.8%	3.6%
Hispanic		3.5%	4.7%	17.3%
White, Not of Hispanic Origin	100%	94.2%	91.4%	33.6%
Multiracial			0.8%	
Students With Disabilities	12.5%	17.4%	11.7%	12.6%
LEP				4.0%
ED	42.5%	19.8%	39.8%	72.4%

Textbook Research and Background

Various research studies show the positive benefits of using a curriculum aligned to National Standards, as well as the benefits of using the NSF-reform textbooks. Research compiled from the Third International Math and Science Study (TIMSS), indicates that a main reason students from top achieving countries outperform students in the U.S. is due to fragmented and incoherent curriculum (Schmidt 1992; McKnight and Schmidt 1998; Schmidt and McKnight 1998; Schmidt, Houang et al. 2004; Schmidt, Wang et al. 2005). Specifically, these authors argue that students experience difficulty learning mathematics and science because topics are introduced, covered briefly, dropped, then brought up again later. Failure of students to master skills due to fragmented coverage prevents students from building a solid foundation of mathematical understanding and ultimately proficiency.

Many of the results from TIMSS helped to inform the processes and methods of research presently conducted for the PROM/SE project. While schools involved with the

PROM/SE project report use of all three NSF-reform textbooks, and the use of others, PROM/SE's focus still remains on the relationship between teaching, content coverage, and curriculum coherence (PROM/SE 2006; PROM/SE 2009; PROM/SE 2009; PROM/SE 2009).

In 2003 a group of researchers from University of Chicago, TERC, and UIC, along with the Consortium for Mathematics and Its Applications reviewed all of the research projects related to NSF-Reform Textbooks and curriculum in their Report of the ARC Center Tri-State Achievement Study (2003). This study reviews all of the impact research collected on the benefits across schools, districts, states, and demographic categories. The research synthesized in this report demonstrates that students using the NSF-reform textbooks consistently outperformed students not using texts aligned to standards. This result held across all of the tests used in the studies, all content-sub-strands, and regardless of SES, gender, and racial/ethnic identity (COMAP, 2003, p. 20). The authors note one caveat; generally, reform students did not outperform comparison students in probability and statistics.

Everyday Mathematics

Everyday Mathematics (EM) is a standards based curriculum designed with two very distinguishing features from traditional mathematics curricula. First, EM is designed with a tiered approach to algorithmic processes. Second, EM uses “distributed practice” to spread the coverage of topics across different mathematical content, with the goal of building a holistic understanding of mathematical operations (UCSMP 2010).

Students first attempt to solve problems without formal algorithmic teaching. Students build on prior experiences to develop algorithms on their own to solve new or novel problems. Students must understand basic number facts and the symbols used in the

problems. Additionally, students must be familiar with the numbering system used by EM. The primary focus in the early grades is on place value and number facts. Students in later grades then use this knowledge to attempt more difficult or complex problems. For example, students asked to subtract one two-digit number from another may use methods like counting up, using manipulatives, or any other invented strategy. Alternative algorithms are presented to assist with formalizing operations, and finally, a third set of Focus algorithms are presented for students not finding success with the first two methods.

The algorithms presented in the EM textbook are different than traditional algorithms designed for pencil-and-paper mathematics computation. For addition, subtraction, multiplication and division, the tiered algorithm process is followed and the actual algorithms are designed to build upon one another. Specifically, multiplication is presented in lattice form, requiring one-digit multiplication, and addition of strings of numbers organized around a lattice. This approach suggests the overall commitment of EM to students first and foremost understanding place value (Bell and Bell 1998-1996). The notation and symbols used can be quite different than those in traditional textbooks.

The second focus of EM is the following of distributed practice. Based on research touting the benefits of a “spaced” versus “massed” practice, content is designed to provide multiple exposure to various important concepts and skills (UCSMP 2000). The distributed nature of the content is designed to ensure that connections are made between different mathematical contexts, and to provide repeated exposure of key ideas.

Investigations

Investigations in Number, Data and Space (INV) was developed at TERC in Cambridge, Massachusetts. The INV curriculum seeks to encourage students to make sense

of mathematical ideas using lessons designed to help students think through new ideas by wrapping them in existing experiences and ideas. Focal points of the INV curriculum include computational fluency with whole numbers, reasoning and sense-making about mathematical ideas, communication skills related to content, and pedagogy to teachers. These points should also help engage all ranges of learning and proficiency in mathematical understanding.

Teachers also play an integral role in the INV curriculum by participating in ongoing learning and professional development opportunities. These opportunities are part of and supported by the INV curriculum. Students usually receive information from teachers, but the INV curriculum specifies that the communication of content and pedagogical strategies also flows from student to teacher. This mutual feedback allows the teachers to make decisions based on these student interactions, while reinforcing the ideas for the students as the ideas are explained and justified (Investigations).

Math Trailblazers

The third NSF reform textbook is *Math Trailblazers* (TB), a project funded by the NSF and based upon research conducted by the University of Illinois, Chicago (UIC), specifically the Teaching Integrated Mathematics and Science Project (TIMS). The TB curriculum focuses on building connections between math and science, with a focused curriculum aligned to the NCTM and State standards. This curriculum encourages children to develop their own problem solving strategies by working in groups, and often discussing possible solutions and the appropriateness of suggested solutions (rather than solving actual problems). This curriculum stresses the importance of students understanding concepts before engaging in “rote” memorization and formal instruction of algorithms (UIC 2003).

A distinctive feature of the TB curriculum, in addition to the encouragement of invented strategies, includes starting with “the tens” rather than starting with “the ones” for addition and subtractions, encouraging grouping and then adding. The curriculum is designed to integrate with science topics in a series of interactive “labs” beginning in the 1st grade and carrying on through 5th grade. Finally, since students are often encouraged to discuss how to solve a problem before they attempt to solve problems, assessments are in the form of rubrics, created to offer partial credit for invented solutions and to encourage group discussions for child-directed learning (UIC 1997, 2004).

Harcourt Math 2004

Although not an NSF reform textbook, according to the publisher’s website, Harcourt Math 2004(HM), is a research-based curriculum designed to align with national standards and is flexible enough to align with individual state standards. HM was designed to “help build conceptual understanding, skill proficiency, problem solving facility, and logical reasoning while carefully developing concepts within and across the mathematics strands” (Houghton, Mifflin et al. 2010).

Methods:

This dissertation focuses on the relationships between curriculum, content coverage, and the patterns of incorrect responses on the mathematics portion of the MEAP. Analysis occurs in three phases. Phase I identifies the common co-occurrences of distractors using a one-mode clustering algorithm to identify the CFGs. Phase II examines the relationships between students using specific textbooks and distractor choices. Phase III involves a statistical analysis of the relationships between textbooks and CFGs which drives the descriptive analysis of the results of the clustering algorithm. The descriptive analysis uses Correspondence Analysis, which is based on the counts of students selecting CFGs based on textbook use, to visually display and confirm the relationships between textbooks and common foils.

Phases I and II – Social Network Models and Correspondence Analysis

Social network models have important potential application for measurement because social network models assume and allow for dependencies among observations (Leinhardt and Wasserman 1979; Holland and Leinhardt 1981). Applying a social network model to test data allows for dependencies in patterns of interactions and thus, patterns in the types of distractors chosen by students. Social network analysis also allows for sparse data. The nature of test items prevents students from choosing more than one option per item and some students select few or no distractors at all. Social network algorithms applied to test data address the issues of data dependencies and sparseness.

A social network is defined by a finite number of nodes and the connections between them. These nodes generally include people, activities, or organizations. The patterns of interactions or relations indicate a flow of activity (Wasserman and Faust

1994). This activity may describe interpersonal communications, common committee memberships, or students choosing from a limited number of multiple choice items.

In this particular study, I employ a social network clustering algorithm developed for studying people and the events they attend (Field, Frank et al. 2006). Rather than analyzing the patterns of behavior represented by people and the events they attend, here the algorithm is used to examine fourth-grade students in nine districts and the distractors chosen on incorrect answers for the mathematics section of the 2008 MEAP. This approach accounts for information that is lost as a result of dichotomizing items by allowing us to look at students and the options they choose simultaneously. Student abilities, background and demographic information, and item information may be added to a predictive model but are not necessary for the clustering of students and distractors.

The Network analysis occurs in two phases. Phase I employs a one-mode analysis focusing on how distractors were chosen together across the sample. Groups of distractors identified by the analyses in Phase I are referred to as Common Foil Groups (CFGs) since these groups represent clusters of distractors (or foils) common across all students in the sample. The model used to identify CFGs is represented by Equation (1).

Equation (1) provides the model for typical one-mode social network data using Frank's (1995) algorithm which exploits the logit model used to estimate the effects of subgroups on social network choices:

$$\log \left| \frac{P[x_{ii'} = 1]}{1 - P[x_{ii'} = 1]} \right| \quad (1)$$

$$= \theta_0 + \theta_1 \text{same subgroup}_{ii'}$$

Where $x_{ii'}=1$ represents a relation between distractors i and i' , 0 otherwise. In this model, θ_0 represents the intercept. θ_1 same subgroup $_{ii'}$ equals 1 if the distractors are in the same subgroup, 0 otherwise. Therefore, θ_1 represents the effect of membership in a common subgroup on the odds of a network nomination. This function, by maximizing θ_1 , allows us to identify cohesive subgroups. Frank (1996) extends the odds ratio to accommodate weighted data with the assumption that the frequency of common choices is measured on an interval scale.

When θ_1 is large, this means actors prefer to interact with members of their subgroups versus members of other subgroups. θ_1 is a network parameter and is insensitive to the marginal probabilities. For the current analysis, when θ_1 is large, distractors in a subgroup are more likely to be chosen together versus distractors from other subgroups. The statistical significance of θ_1 is determined using a likelihood ratio test between the following models:

$$\log P(X_{ij} = x_{ij}) = \quad (2)$$

$$\theta_0 + \theta_{1base} \text{same group}_{ii'}$$

and

$$\log P(X_{ij} = x_{ij}) \quad (3)$$

$$= \theta_0 + \theta_{1base} \text{same group}_{ii'}$$

$$+ \theta_{1subgroup\ processes} \text{same group}_{ii'}$$

Phase II uses a two-mode analysis looking at how individual students' chosen distractors are grouped by districts and textbooks. Groups of students and distractors identified in the two-mode analysis are referred to as Textbook Foil Groups (TFGs) and represent the commonly chosen distractors for students using particular textbooks. The model used to identify TFGs is represented by Equation (4).

The Field et al. algorithm (2006) identifies clusters by maximizing the odds of an actor participating in an event, or in our case, a student selecting a particular distractor option inside his cluster and relative to events outside of his cluster. The clusters identified ultimately consist of actors and events simultaneously, without reducing the affiliation data to a single mode.

The objective function employed by Field et al. in Equation (4) is similar to that derived from Equation (1), but allows for two-mode data rather than a single mode. Typically, in social network analysis, one mode data analysis is appropriate for actors nominating other actors, while two-mode data analysis looks at people attending events, for example. Here, I use two-mode analysis since these data represent students choosing distractors. The logit model for two-mode data is defined as $y_{ij} = 1$ if student i chooses distractor choice j , 0 otherwise.

$$\log\left(\frac{p[y_{ij} = 1]}{1 - p[y_{ij} = 1]}\right) = \theta_0^* + \theta_1^* \text{same cluster}_{ij} \quad (4)$$

Here θ_1^* same subgroup $_{ij}$ equals 1 if student i is assigned to the same cluster as distractor j , 0 otherwise. Thus, θ_1^* is large whenever students choose distractors within their cluster and when students do not choose distractors outside of their position.

Equivalently, using Table 2.4 θ_1^* is large for AD large and CB small. Using the ratio, between equations (5) and (6), a change in X^2 , enabling testing of the statistical significance of the TFGs.

$$\log P(X_{ij} = x_{ij}) = \theta_0 + \theta_{1base} \text{same position}_{ij} \quad (5)$$

$$\begin{aligned} \log P(X_{ij} = x_{ij}) \\ &= \theta_0 + \theta_{1base} \text{same position}_{ij} \\ &\quad + \theta_{1subgroup \text{ processes}} \text{same position}_{ij} \end{aligned} \quad (6)$$

Table 2.4 : Odds Ratio Table for Position Membership and Event Participation

	Distractor Chosen	
Position Membership	No [$y_{ij} = 0$]	Yes [$y_{ij} = 1$]
Different, 0	A	B
Same, 1	C	D

This approach allows us to look at actors and events simultaneously without having to reduce the affiliation or event data into a uni-mode dataset (Wasserman & Faust, 1994), while not assuming observations are independent. I will apply the algorithm to examine students and their pattern of incorrect options simultaneously, while allowing for dependencies in these response patterns among items and students.

Once the TFG positions were assigned to the students and items from each school, I remerged these data with the demographic and test data. Summary statistics were computed and reported in the appendix. Statistics produced by the KlugeFinder software

include transformed z-scores for the theta values, along with corresponding p-values. These p-values help to determine whether or not θ_1 or θ_1^* , respectively, are statistically different than zero. If θ_1 or θ_1^* are statistically different than zero, this suggests that the clusters of distractors and/or students and distractors are not likely to occur at random.

Correspondence Analysis

The next step employs Correspondence Analysis (CA) to examine the relationship between textbook and Foil Groups; specifically, CA provides a metric of each CFG and textbook's contribution to the overall Chi-Square distance as well as graphic representations of proximities of CFGs to textbooks. The Chi-Square distance is used to measure and depict the distances between profile points. The distances are weighed, but weights are assigned to dimensions rather than to the data or profile points in the analysis. The distance between two rows i and i' is given by

$$d^2(i, i') = \sum_{j=1}^J \frac{1}{n_{+j}} \left(\frac{n_{ij}}{n_{i+}} + \frac{n_{i'j}}{n_{i'+}} \right)^2 \quad (7)$$

The distance between two columns j and j' is given by

$$d^2(j, j') = \sum_{i=1}^I \frac{1}{n_{i+}} \left(\frac{n_{ij}}{n_{+j}} + \frac{n_{ij'}}{n_{+j'}} \right)^2 \quad (8)$$

Each square is weighted by the inverse of the frequency corresponding to each term.

The Chi-Squared distance is a yardstick to measure dissimilarities among points. CA is related to principal components analysis and is an exploratory technique designed to find multidimensional representation of the association between the rows and columns of a

two-way contingency table. While principal components uses a constant, Euclidean metric, CA uses expected abundances as a metric. Expected abundances from marginal totals are used in the same way a chi-square analysis uses contingency tables. This technique identifies scores for the row and column categories on a small number of dimensions which account for the greatest proportion of the chi-square for association between the row and column categories.

The measures used in the CA include a frequency table containing the number of times a specific distractor was chosen from a specific CFG by a student using a particular textbook. The CA includes the CFGs and textbooks, as the TFGs are already determined by textbook. The results of the analysis illuminate the specific connections between certain textbooks and CFGs and provide a comprehensive interpretation of the CFGs and TFGs.

Using a two-way contingency table with rows representing CFGs and columns representing textbooks, this CA analysis is designed to show how the data deviated from expected values when the CFGs and textbooks are independent. For this two-way table the scores from the CFGs, say x_{im} , and textbooks, y_{jm} , on dimension $m = 1, \dots, M$ are derived from a singular value decomposition of residuals from independence expressed as $d_{ij} / \sqrt{n}$,

to account for the largest proportion of the chi-square in a small number of dimensions. Each cell within the contingency table contributes to the chi-square statistic value. Per Friendly (1991, p.514), the matrix of deviations from independence is expressed in Equation (9).

$$\frac{d_{ij}}{\sqrt{n}} = \sum_{m=1}^M \lambda_m x_{im} y_{jm} \quad (9)$$

where $\lambda_1 > \lambda_2 \geq \dots > \lambda_M$, and $M = \min(I - 1, J - 1)$. If there are M dimensions, the decomposition in Equation (9) is exact. A rank- d approximation in d dimensions is obtained from the first d terms on the right side of (9). The sums across the λ_M values are equal to the mean squared contingency coefficient and decomposed into the linear components of CA.

Phase III –Teacher Consultations, Factor Analysis, and BCLM Logistic Regression

Following Phases I and II of the network analysis, Phase III begins by combining the attribute codes assigned to each item and distractor by teacher experts with the outcomes of the CFG and TFG analysis. The goals of Phase III include: identifying (beyond content) the attributes of distractors making up each CFG and comparing the composition of the CFGs based on these attributes; identifying the relationship between CFGs and TFGs and/or CFGs and Textbooks; determining the statistical significance of CFGs and TFGs and how this relates to teacher experiences and literature reviews.

Teacher Consultations

Prior to the analysis, I trained three teachers to assign process and skill codes to each of the items, following the methods of Tatsuoka et al. (2004). I included two additional teachers to assist in analyzing the results. The teachers included two teachers from Illinois, one teacher from Arizona, and two teachers from Michigan. For ease of discussion, each teacher will be assigned a one-letter pseudonym. Both teachers from

Illinois have used the EM curriculum for over five years and have participated in professional development programs at the University of Chicago to help them fully understand and implement the curriculum. Both Illinois teachers have been teaching for seven years at the elementary level (Ms. E. and Ms. J.). The teacher from Arizona has been using INV for the past three years (Mrs. R.). Her professional development focused not specifically on the INV curriculum, but rather on teaching mathematics to LEP students. She has been teaching elementary mathematics for 25 years. One of the teachers from Michigan has been using the TB curriculum since she began teaching seven years ago (Mrs K). The second teacher from Michigan has used EM for the last two years, and various other materials before her district adopted the EM curriculum (Mrs M.). She has been teaching elementary school mathematics and reading for 22 years.

The three teachers were enlisted to code the items, sign confidentiality agreements related to the item content, and file these agreements with the State of Michigan's Office of Educational Assessment & Accountability. The coders included Ms E., Ms. J., and Mrs. K. . In order to compensate these teachers for their time and expertise, each teacher received a \$100 gift certificate from Barnes & Noble (\$50 for the item coding and \$50 for the analysis). Mrs. R and Mrs. M. both received \$50 gift certificates for their time in helping to analyze the results.

I assigned 16 items to each of the three coders, with two items coded by all three teachers and then six items shared two-by-two by all three teachers. I coded all of the items and then analyzed the degree of agreement between the coders and the ten process and ten skill categories. The percentage of agreement was 94% for the process codes and 89% for the skill codes. For items where there were differing categories chosen between

teachers, I asked all of the teachers to review the items again and, as a group, to come to an agreement on the appropriate codes. This was facilitated via conference calls and electronic mail. The final analysis uses the revised codes and lead to 100% agreement on both scales.

After completing the analysis and identifying the positions for students and items in each school and by curriculum, I asked the teachers to look at the results and explain how their experience with teaching students and student mistakes related to the patterns identified by the algorithm. The teacher explanations are combined with existing literature on commonly known cognitively correlated math skills and item types as well as differences in how material is presented depending on the textbook.

Principal Factor Analysis and Logistic Regression

The first step in understanding the nature of the CFGs is to understand the attributes of the distractors making up each CFG and how they vary across CFGs. The second step is to examine these effects while controlling for textbook. These analyses illuminate the nature of the CFGs, individually and with respect to textbooks. Further, these analyses explain the relationships between the TFGs and the CFGs.

Factor Analysis

In order to examine the effects of the item attributes on the odds of CFG membership, I used the results of a principal factor analysis (PFA) to determine the structure of the relationships between each of the item attributes. I set the prior communality estimate for each variable to its squared multiple correlation with all other variables, thus enabling computations of the principal factors (rather than principal

components). The rotation method for this analysis is Promax as the assumptions of an oblique rotation fit the nature and assumptions around the data more accurately than would an orthogonal rotation.

The factors identified in this analysis represent the underlying skill or process deficiencies associated with the distractors chosen regardless of CFG or textbook. The factors identified using PFA mimic the underlying attributes identified by teachers. These factors also help to explain the relationships among the CFGs as well as the relationships between CFGs and Textbooks.

Multinomial Baseline-Category Logit Model

The general model for logistic regression, presented in Equation (10), may be used when trying to measure the effect of binomial indicator variables on classification of observations into categories. Here, $\mathbf{x}$ is a vector of explanatory variables and $\pi = P(Y = 1 | \mathbf{x})$ is the response modeled. Additionally α is the intercept parameter while $\boldsymbol{\beta}$ is the vector of slope parameters.

$$\text{logit}(\pi) \equiv \left(\frac{\pi}{1 - \pi} \right) = \alpha + \boldsymbol{\beta}'\mathbf{x} \quad (10)$$

For these analyses, the desired model has CFG classification as the response probability to be modeled, and the explanatory variables are ten process attributes and ten skill attributes. In order to model categorical outcomes, I use a generalized or baseline-category logit model (BCLM). The BCLM calculates the odds of CFG membership, holding CFG 1 as a reference group. Equation (11) represents the BCLM. Here, $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_k$ are 12

CFG intercept parameters and $\beta_1, \dots, \beta_k$ are the k vectors of the slope parameters. Here the response, Y , is restricted to the $1, \dots, k, k+1$ values. Since there are 13 CFGs, k here is 12.

$$\log \left(\frac{P(Y = 1|\mathbf{x})}{P(Y = k + 1|\mathbf{x})} \right) = \alpha_i + \beta'_i \mathbf{x}, \quad i = 1, \dots, k \quad (11)$$

Since the outcome CFG is nominal and there is no “natural” ordering to these groups, Equation (11) is the most appropriate. This model is a case of the discrete choice or conditional logit models (SAS technical information, McFadden 1974).

Measures

Social Network Analysis

The variable of interest in Social Network Analysis generally reflects some measure of association between two nodes (i.e. actors or events). In these analyses, the one-mode analysis uses the count of times each pair of distractors is chosen together. The two-mode analysis uses a binary indicator for a student selecting a distractor; the weight equals “1” if a student chose a given distractor, “0” otherwise.

The CliqueFinder software, used to run the Field et al. algorithm, requires a list file with three columns (To, From, Weight). For the CFG (one-mode) analysis, one file containing item responses across the entire sample is used. For the TFG (two-mode) analysis, a separate file is needed for each school. Since each distractor represents a unique mistake or misconception, each required a unique ID. If a student misses five items, she will have five lines in this file, one for each of the incorrect distractors chosen. Since each item only appears once, the weight for each entry is always one. An example of the data for the one-mode analysis, one distractor would look like that shown in Figure 2.1

Figure 2.1: Example of One-Mode Data

Distractor ID (to)	Distractor ID (from)	Weight
03B	19A	10
03B	22A	2
03B	33B	1
03B	56A	36

This list indicates that distractor “03B” was chosen in common with the distractor “19A” 10 times, two times with distractor 22A, and so on. The weights indicate number of times each pair of distractors was mutually selected across the entire sample. The maximum value for the weight is the sample size.

The list file for the two-mode data is slightly different. An example of the data for the two-mode analysis, one student would look like Figure 2.2:

Figure 2.2: Example of Two-Mode Data

Student ID (to)	Distractor ID (from)	Weight
99999	19A	1
99999	22A	1
99999	33B	1
99999	56A	1

This list indicates that Student 99999 chose distractors “19A”, “22A”, “33B”, and “56A”. Since each item can only be answered once, the maximum weight for any distractor is one.

Existing literature suggests omitting the correct responses and only looking at the patterns of incorrect choices. Green et al. looked only at the incorrect options, as including the correct responses tends to confound the results and may mask the subtleties of the

differences in distractor choices. For this analysis, only the incorrect responses are analyzed¹. The differences in clustering, resulting from the inclusion or not of correct and omitted responses, follow the findings of Green, Crone, and Folk (1989), which suggest no significant differences in the outcomes or interpretation of the results.

Principal Factor Analysis and Logistic Regression

Distractor-attribute codes for the processes and skills necessary to answer an item correctly were merged with a dataset containing the following: all of the distractor vectors for each student in the sample, CFG information by distractor, TFG information for a student and a distractor, and the textbook. Attribute codes consisted of ten process attributes and ten skill attributes. These attributes are listed in Table 2.5. Skill type S10 is crossed out since there were no open-ended items in this assessment. In nominal response logistic models, where the $k+1$ possible responses have no natural ordering, the general logit model is extended to a multinomial model known as the baseline-category logit model (BCLM). The measures for this analysis include counts of attribute codes, by CFG with and without controlling for textbook use.

¹ I replicated the analysis including correct and missing responses. The analysis including the correct responses results in one additional group, containing all of the correct choices. The analysis including missing responses randomly allocates the missing responses to groups. The interpretation is the same for all analyses, and thus the most parsimonious model remains.

Table 2.5 : Item and Distractor Attribute Codes

Code	Attribute_Strand	Attribute
P1	Process	Translate/formulate equations and expressions to solve a problem
P2	Process	Computational applications of knowledge in arithmetic and geometry
P3	Process	Judgmental applications of knowledge in arithmetic and geometry
P4	Process	Applying rules in algebra
P5	Process	Logical reasoning - includes case reasoning, deductive thinking skills, if-then, necessary and sufficient, generalization skills
P6	Process	Problem search: analytic thinking, problem restructuring; inductive thinking
P7	Process	Generating, visualizing, and reading figures and graphs
P8	Process	Applying and evaluating mathematical correctness
P9	Process	Management of data and procedures
P10	Process	Quantitative and logical reading
S1	Skill (item type)	Unit conversion
S2	Skill (item type)	Apply number properties and relationships; number sense/number line
S3	Skill (item type)	Using figures, tables, charts and graphs
S4	Skill (item type)	Approximation/estimation
S5	Skill (item type)	Evaluate/verify/check options
S6	Skill (item type)	Patterns and relationships (inductive reasoning skills)
S7	Skill (item type)	Using proportional reasoning
S8	Skill (item type)	Solving novel or unfamiliar problems
S9	Skill (item type)	Comparison of two/or more entities
S10	Skill (item type)	Open-ended items, in which an answer is not given
S11	Skill (item type)	Understanding verbally posed questions

The same dataset is used for the PFA, BCLM, and CA. This dataset has a record for each student, and contains the CFG and TFG identifiers, distractor response strings, school level information (including textbook), and background demographic information. The PFA utilizes the counts of attribute codes associated with each distractor chosen for each

student. The factor scores are merged with the original dataset enabling the BCLM analysis. The BCLM analysis uses the counts of the specific distractors chosen by students, controlling for textbooks, based on factor scores. For the CA, the counts of students and items by textbook and CFG, respectively, were used to analyze and graphically display the final results.

CHAPTER 3 : ANALYSIS

Phase I – Common Foil Groups

The first Phase of this analysis examines the co-occurrences of distractor choices across the entire sample of students. The network analysis identifies the CFGs and parameter estimates for the network statistics. Recall from Chapter 2, the logit models for both one- and two-mode data are defined, respectfully, as (1) and (4). In Equation (1), $x_{ii'} = 1$ represents a relation between distractors i and i' , and is equal to 0 otherwise. Here, θ_0 represents the intercept and θ_1 same subgroup ii' equals 1 if distractors i and i' are in the same subgroup and is equal to 0 otherwise. Therefore, θ_1 represents the effect of membership in a common subgroup on the odds of a network nomination.

The focal point of this analysis is on the effect of subgroup membership on the odds of two distractors being chosen together by a student. Specifically, what are the odds of two distractors being selected together when they come from the same subgroup relative to distractors outside of the same subgroup?

The Cliquefinder software produces files indicating the subgroup membership for each of the distractors as well as parameter estimates and statistical tests of these parameters. Table 3.1 provides the parameter estimates generated by the Cliquefinder software. The predicted value is for unweighted data and assumes that the tendency for weights to be concentrated within the CFGs is not greater than the tendency for the presence of ties. The likelihood ratio test uses the ratio of Equation (2)/Equation(3).

Table 3.1: Odds Ratio and Likelihood Ratio test for Common Foil Groups

	Odds Ratio	Log Odds	(Log Odds/2)	θ_1 Subgroup Processes	Approx LRT	P-value
Observed	6.24	1.83	0.91	0.07	92.67	0.00

As the results in Table 3.1 indicate, we can reject the null hypothesis that items are chosen together by chance alone. Based on these results it is appropriate to examine the nature of the distractors within each subgroup defined as the CFGs. The one-mode analysis identified 38 CFGs each containing unique clusters of distractor choices. Of these 38 groups, only 13 contained at least two percent of distractors chosen; these 13 remaining CFGs are retained for the analysis.

Table 3.2 contains the results of the one-mode Cliquefinder analysis as well as classical item statistics at the distractor level. CFG 1 contains the largest number of distractors (26) and is the only CFG where there are multiple distractors for an item appearing in the same group. This group represents the distractors most commonly chosen across all students and textbooks. The proportion of correct answers for the items (item-level p-values) in CFG 1 is 0.44 and is the lowest of all CFGs. The proportions of students choosing these distractors are higher than any other CFG and also represent the mistakes most commonly made for the upper and lower thirds of the score distribution. Almost 50% of distractors chosen come from CFG 1; this is true when looking at the percentages of distractors chosen from CFG 1 by textbook as well.

Table 3.2: Item and Distractor Information by Common Foil Group

Distractor	Group	Item Description	Distractor Description	Item p-value	Distractor p-value	Distractor Point Biserial	Key Point Biserial
a08_B	1	Estimate sum / difference of two 3-digit numbers	Chose option closest to both values rather than the sum.	0.594	0.175	-0.330	0.536
a08_C	1	Estimate sum / difference of two 3-digit numbers	Rounded 585 to 500 and underestimated	0.594	0.181	-0.294	0.536
a12_D	1	Recognize multiplication and division situations	Chose - over ÷ as the correct operation	0.529	0.215	-0.141	0.484
a15_A	1	Understand meaning & terminology of fractions	Chose reciprocal of correct answer	0.520	0.398	-0.228	0.360
a16_B	1	Understand meaning & terminology of fractions	Confused numerator and denominator	0.567	0.395	-0.167	0.240
a17_A	1	Know benchmark temperatures & compare cooler, warmer	Chose 0°F as freezing point of water	0.567	0.385	-0.325	0.411
a18_A	1	Know benchmark temperatures & compare cooler, warmer	Chose 112°F as the boiling point of water	0.468	0.198	0.003	0.098
a19_C	1	Compose and decompose triangles and rectangles	Did not combine or rotate shapes properly	0.424	0.196	-0.132	0.331
a19_D	1	Compose and decompose triangles and rectangles	Combined but did not rotate shapes	0.424	0.213	-0.080	0.331
a22_D	1	Identify, describe, classify familiar 3-D solids	Confused definitions for prism and pyramid	0.484	0.272	-0.142	0.292

Table 3.2 (cont'd)

Distractor	Group	Item Description	Distractor Description	Item p-value	Distractor p-value	Distractor Point Biserial	Key Point Biserial
a23_C	1	Read scales on axes. Identify the max, min, range	Confused max with min	0.493	0.398	-0.359	0.409
a24_A	1	Read scales on axes. Identify the max, min, range	Chose max rather than the range of data on a graph	0.190	0.420	-0.092	0.313
a24_B	1	Read scales on axes. Identify the max, min, range	Chose values of first and last columns on graph, not range	0.190	0.235	-0.071	0.313
a26_C	1	Recognize, name and use equivalent fractions	Chose 4/4 rather than $4/8=1/2$	0.256	0.241	-0.178	0.469
a26_D	1	Recognize, name and use equivalent fractions	Chose reciprocal of correct answer	0.256	0.444	-0.153	0.469
a27_B	1	Model +, - of fractions on number line	Chose + over - as the correct operation	0.326	0.281	-0.087	0.209
a27_C	1	Model +, - of fractions on number line	Did not recognize subtraction necessary from diagram	0.326	0.260	-0.016	0.209
a29_C	1	Identify points, line segments, lines and distance	Chose line segment instead of a line	0.535	0.443	-0.160	0.217
a31_C	1	Read & interpret horizontal and vertical bar graphs	Confused horizontal and vertical axis	0.615	0.213	-0.271	0.341

Table 3.2 (cont'd)

Distractor	Group	Item Description	Distractor Description	Item p-value	Distractor p-value	Distractor Point Biserial	Key Point Biserial
a34_C	1	Understand meaning of 0.50 & 0.25 related to money	Didn't read stem completely and divided by 3 rather than 4 Chose 1 hour 15 minutes as duration from 1:15 p.m. to 2:45 p.m.	0.470	0.257	-0.015	0.317
a38_C	1	Measure in mixed units within measurement system		0.659	0.204	-0.317	0.469
a40_C	1	Use relationships between sizes of standard units	Chose 250 cm > 25 m	0.531	0.245	-0.230	0.442
a42_A	1	Calculate area and perimeter of square & rectangle	Added $l+w$ to find perimeter rather than $2(l+w)$	0.370	0.487	-0.354	0.434
a46_C	1	Solve problems using bar graphs, compare graphs	Misread bar graph by one unit	0.657	0.221	-0.063	0.359
a48_A	1	Solve problems about perimeter/area of rectangles	Added $l+w$ to find area rather than $l \times w$	0.168	0.237	-0.333	0.304
a48_B	1	Solve problems about perimeter/area of rectangles	Calculated perimeter rather than area	0.168	0.558	0.107	0.304
a32_B	3	Identify operation for problem and solve	Chose $\times$ over $+$ as the correct operation	0.795	0.172	-0.279	0.521

Table 3.2(cont'd)

Distractor	Group	Item Description	Distractor Description	Item p-value	Distractor p-value	Distractor Point Biserial	Key Point Biserial
a33_D	3	Identify operation for problem and solve	Chose + over - as the correct operation	0.722	0.131	-0.205	0.518
a35_A	3	Understand meaning of 0.50 & 0.25 related to money	Chose 1/2 of \$5.00=\$2.15	0.702	0.084	-0.256	0.448
a39_B	3	Measure in mixed units within measurement system	Chose 15 yards = 5 feet	0.614	0.169	-0.225	0.375
a21_A	4	Identify, describe, classify familiar 3-D solids	Identified a cone as a triangular pyramid	0.833	0.106	-0.292	0.237
a35_B	4	Understand meaning of 0.50 & 0.25 related to money	Chose 1/2 of \$5.00=\$2.25	0.702	0.110	-0.257	0.448
a49_B	4	Find solutions to open sentences that use x and ÷	Chose 7 to complete the open sentence: $9 \times \square = 72$	0.517	0.212	-0.213	0.164
a11_D	10	Recognize multiplication and division situations	Chose $40/5=9$	0.785	0.098	-0.356	0.431
a19_A	10	Compose and decompose triangles and rectangles	Did not combine or rotate shapes properly Chose 6,931 as a value	0.424	0.156	-0.214	0.331
a25_A	10	Compare and order numbers up to 10,000.	between 5,642 and 6,633	0.573	0.138	-0.192	0.473

Table 3.2 (cont'd)

Distractor	Group	Item Description	Distractor Description	Item p-value	Distractor p-value	Distractor Point Biserial	Key Point Biserial
a45_C	10	Add and subtract money in dollars and cents	Chose \$20.00-\$17.25=\$3.25	0.542	0.161	-0.327	0.323
a05_C	11	Add and subtract thru 999 w/regrouping, 9,999 w/o	Chose 76-29=53	0.810	0.130	-0.163	0.358
a35_D	11	Understand meaning of 0.50 & 0.25 related to money	Chose 1/2 of \$5.00=\$2.75	0.702	0.100	-0.158	0.448
a40_D	11	Use relationships between sizes of standard units	Chose 25 centimeters = 25 meters	0.531	0.101	-0.248	0.442
a12_A	13	Recognize multiplication and division situations	Chose - over ÷ as the correct operation	0.529	0.131	-0.141	0.484
a24_C	13	Read scales on axes. Identify the max, min, range	Chose the min rather than the range of the data on a graph	0.190	0.151	-0.187	0.313
a25_B	13	Compare and order numbers up to 10,000.	Chose 5,610 as a value between 5,642 and 6,633	0.573	0.150	-0.132	0.473
a43_A	13	Calculate area and perimeter of square & rectangle	Chose 30 as perimeter for a square with sides measuring 5 inches	0.622	0.126	-0.506	0.439
a12_C	14	Recognize multiplication and division situations	Chose + over ÷ as the correct operation	0.529	0.123	-0.437	0.484

Table 3.2 (cont'd)

Distractor	Group	Item Description	Distractor Description	Item p-value	Distractor p-value	Distractor Point Biserial	Key Point Biserial
a25_C	14	Compare and order numbers up to 10,000.	Chose 6,745 as a value between 5,642 and 6,633	0.573	0.135	-0.251	0.473
a34_B	14	Understand meaning of 0.50 & 0.25 related to money	Chose \$2.00 divided between 4 people evenly equals \$1.00	0.470	0.121	-0.238	0.317
a43_D	14	Calculate area and perimeter of square & rectangle	Added $l+w$ to find perimeter rather than $2(l+w)$	0.622	0.126	-0.132	0.439
a13_C	15	Find products to 10 X 10 and related quotients	Chose $8 \times 7 = 54$	0.758	0.095	-0.179	0.416
a20_A	15	Compose and decompose triangles and rectangles	Decomposed shapes incorrectly and did not rotate	0.650	0.121	0.042	0.335
a34_A	15	Understand meaning of 0.50 & 0.25 related to money	Multiplied $3 \times \$2.00$ instead of dividing \$2.00 by 4	0.470	0.147	-0.040	0.317
a27_D	16	Model +, - of fractions on number line	Chose $1-5/10$ rather than $8/10-5/10$ from diagram	0.326	0.127	-0.176	0.209
a30_C	16	Identify, describe, compare, classify 2-D shapes	Chose a triangle as a possible side view of a rectangular prism	0.685	0.143	-0.165	0.378

Table 3.2 (cont'd)

Distractor	Group	Item Description	Distractor Description	Item p-value	Distractor p-value	Distractor Point Biserial	Key Point Biserial
a39_C	16	Measure in mixed units within measurement system	Chose 1 yard 2 feet = 5 feet	0.614	0.128	-0.102	0.375
a45_D	16	Add and subtract money in dollars and cents	Chose \$20.00-\$17.25=\$3.75	0.542	0.162	-0.186	0.323
a22_B	17	Identify, describe, classify familiar 3-D solids	Chose triangular prism over rectangular prism	0.484	0.119	-0.120	0.292
a30_D	17	Identify, describe, compare, classify 2-D shapes	Chose a pentagon as a possible side view of a rectangular prism	0.685	0.158	-0.197	0.378
a31_A	17	Read & interpret horizontal and vertical bar graphs	Confused and misread axis labels and values	0.615	0.102	-0.119	0.341
a18_B	18	Know benchmark temperatures & compare cooler, warmer	Chose 180° as the boiling point of water	0.468	0.171	-0.154	0.098
a20_B	18	Compose and decompose triangles and rectangles	Decomposed shapes but did not rotate	0.650	0.162	-0.215	0.335
a40_B	18	Use relationships between sizes of standard units	Chose 25 meters < 25 centimeters	0.531	0.119	-0.157	0.442

Table 3.2 (cont'd)

Distractor	Group	Item Description	Distractor Description	Item p-value	Distractor p-value	Distractor Point Biserial	Key Point Biserial
a49_D	18	Find solutions to open sentences that use x and ÷	Chose 9 to complete the open sentence: $9 \times \square = 72$	0.517	0.177	-0.134	0.164
a28_B	19	Distinguish between units of length and area in cont	Chose square inches to represent length	0.784	0.110	-0.130	0.312
a33_B	19	Identify operation for problem and solve	Didn't borrow for subtraction; chose $428 - 386 = 142$	0.722	0.087	-0.244	0.518
a36_D	19	Use common measures of length, weight, time	Didn't switch between a.m. and p.m. properly	0.700	0.142	-0.061	0.405
a18_C	38	Know benchmark temperatures & compare cooler, warmer	Chose 200° as the boiling point of water	0.468	0.161	-0.087	0.098
a37_B	38	Use common measures of length, weight, time	Chose kilogram as the correct unit of measure to measure distance	0.709	0.146	-0.215	0.416
a43_B	38	Calculate area and perimeter of square & rectangle	Chose 25 as perimeter for a square with sides measuring 5 inches	0.622	0.122	-0.189	0.439
a45_A	38	Add and subtract money in dollars and cents	Chose \$20.00 - \$17.25 = \$2.25	0.542	0.133	-0.077	0.323

Phase II – Textbook Foil Groups and Correspondence Analysis

Phase II explores the relationships between the specific textbooks students use and the distractors they choose. This two-mode (student-to-distractor) analysis identifies co-occurrences of students and distractors that are specific to a textbook. As with the one-mode analysis, parameter estimates are produced via the Kluefinder software along with test statistics enabling evaluation of a null hypothesis of $\theta_1^* = 0$.

Before interpreting the two-mode results, I examined the values of $\theta_{1 \text{ subgroup processes}}$ compared to the values of $\theta_{1 \text{ base}}^*$ based on simulations built into the software. Kluefinder computes a likelihood ratio test between the models represented by Equation (5) and Equation (6).

Table 3.3 provides the observed odds ratio, the log of the odds ratio, the log of the odds ratio divided by two (the total value of θ_1), and the value of $\theta_{1 \text{ subgroup processes}}$, along with the respective z-scores and p-values. Small p-values indicate that we may reject the null hypothesis that $\theta_{1 \text{ subgroup processes}}$ is equal to zero, thus providing evidence that students choose distractors within TFGs at a rate that is unlikely to have occurred by chance alone.

Table 3.3: Odds Ratio and Likelihood Ratio test for Common Foil Groups

Curriculum	School	Odds Ratio	Log Odds	(Log Odds/2)	θ_1	Z-Score	P-value
					Subgroup Processes		
Everyday Math	EM1	6.23	1.83	0.91	0.46	6.66	0.00
	EM 2	6.74	1.91	0.95	0.25	10.24	0.00
	EM3	6.84	1.92	0.96	0.06	14.91	0.00
	EM4	7.05	1.95	0.97	0.55	12.83	0.00
Investigations	INV 1	5.57	1.72	0.86	0.49	5.48	0.00
	INV 2	5.11	1.63	0.82	0.59	3.30	0.00
	INV 3	7.25	1.98	0.99	0.40	8.68	0.00
Trailblazers	TB1	5.99	1.71	0.85	0.43	6.84	0.00
	TB2	5.51	1.44	0.72	0.65	5.96	0.00
	TB3	6.01	1.79	0.90	0.56	8.53	0.00
HM	HM1	6.57	1.88	0.94	0.50	6.49	0.00
	HM2	6.33	1.85	0.92	0.53	7.01	0.00
Mixed Texts	MIX 1	4.37	1.47	0.74	0.64	1.47	0.50
	MIX 2	2.01	0.70	0.35	0.23	1.68	0.50
	MIX3	3.72	1.31	0.65	0.22	1.14	0.50
	MIS4	1.56	0.44	0.22	0.35	1.58	0.50

The results in Table 3.3 indicate that we may reject the null hypothesis of subgroup processes equaling zero for the majority of districts reporting the use of one common textbook. These results do not hold for districts reporting the use of multiple textbooks across schools. For the districts reporting the use of multiple textbooks, it is not clear if the TFGs are anything but random noise. TFGs are inconsistent across districts reporting multiple texts; a different result than districts reporting a common textbook across schools. For single-textbook districts with statistically insignificant TFGs, the TFGs themselves are consistent across districts using the same textbook.

Results from the network analysis indicate that while we may retain Hypothesis I for the CFGs, the results for the TFGs indicate that we should reject Hypothesis I for some schools and districts, and retain it for others. The results from the two-mode analysis are

interesting regardless of the statistical significance, as the actual TFGs identified are consistent across schools and vary only in the level of significance. The sizes of these schools vary within textbook classifications, and for smaller schools and districts the results may be too sparse to estimate fully.

Correspondence Analysis

The TFGs identified in the two-mode analysis are very similar to the CFGs identified in the one-mode analysis and the variations reflect differences in the textbook and curriculum to which a student is exposed in the classroom. The results from the CA provide a visual display of the relationships between CFGs and textbooks. These results help to understand how different textbooks are associated with different types of distractor choices as well as informing the Factor Analysis and BCLM analyses.

Phase III – Principal Factor Analysis and BCLM Logistic Regression

In Phases I and II, the null of Hypothesis I was rejected for CFGs. However, this is not true for all TFGs. The co-occurrences of distractor choices are not likely random. The co-occurrences of students choosing distractors follow the general patterns observed in the CFGs, but with differences by textbook and less certainty about the strength of the TFGs.

Prior to fitting any models, I merged the 20 item attribute codes with the student records containing the vector of distractor responses as well as textbook, CFG, and school and district identifying information. This dataset serves the needs for both Factor Analysis and BCLM.

The results from the one-mode analysis suggest that the co-occurrences of distractors are not likely to occur by chance alone, but there is a need to make sense of the

underlying reasons behind different distractors having a larger likelihood of being chosen together. Looking at the types of mistakes and the content, the CFGs seem to be composed of varying content, but with related mistakes. The TFGs indicate that there are distinctions in the way different textbooks influence the types of mistakes students are likely to make, and that there is an overall pattern to how distractors are likely to be chosen.

If we suppose that the CFGs represent the global set of distractor groupings, and that the TFGs represent the same groups while controlling for textbook, then it is logical to ask what impact item attributes and textbook exposure have on CFG classification. Specifically, how do textbooks and item attributes affect the odds of distractors being classified into the groups defined by the one-mode KF algorithm?

Factor Analysis of Process and Skill Codes

In order to examine the effects of the item attributes on the odds of CFG membership, I used the results of a principal factor analysis (PFA) to determine the structure of the relationships between each of the item attributes. Using the options within the SAS software, the model used here sets the prior communality estimate for each variable to its squared multiple correlation with all other variables, thus enabling computations of the principal factors (rather than principal components). I use an oblique Promax rotation to obtain the final factor structures as the assumptions of an oblique rotation fit the nature and assumptions around the data more accurately than would an orthogonal rotation. Within the SAS PFA software, STRATA variables may be defined to provide a hierarchical nature to the analysis. This is appropriate since the distractors are nested within CFGs. These factors are defined, then, based on the CFGs.

When assessing the appropriateness of a common factor model, I compared the differences between the partial correlations to the original correlations to verify that the partials (controlling for all other variables) were small compared to the original correlations. I also examined Kaiser's measure of sampling adequacy (MSA) and was able to identify the appropriate attribute codes for the common factor model. Eleven attribute codes were used in the final PFA analysis.

BCLM Logistic Regression

After merging the item attributes with the student data, I fit a BCLM using logistic regression to determine: (1) The odds of an item being classified in a particular CFG based on Textbook and (2) The odds of factor association based on Textbook. The outcome from (1) explains the direct relationship between textbooks and CFGs. The outcome from (2) explains how the types of mistakes are likely to be impacted by a student's curriculum.

Model (1) CFGs and Textbooks

The first model examines the odds of CFG classification based solely on the textbook a student used. This model is shown in Equation (12).

$$\log \left(\frac{P(Y = 1|\mathbf{x})}{P(Y = k + 1|\mathbf{x})} \right) = \boldsymbol{\alpha}_i + \boldsymbol{\beta}'_i \mathbf{x}, \quad i = 1, \dots, k \quad (12)$$

where

$\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_k$ are the k vector of intercepts for

CFGs

$\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_k$ are the k vectors of the slope

parameters

x_1 =Everyday Math

x_2 =Investigations

x_3 =Trailblazers

x_4 =Harcourt Math 2004

Model (2) Textbooks and Factors

The second model examines the odds of factor association, by textbook use based on the counts of each factor across the CFGs. This model is shown in Equation (13).

$$\log \left(\frac{P(Y = 1|\mathbf{x})}{P(Y = k + 1|\mathbf{x})} \right) = \boldsymbol{\alpha}_i + \boldsymbol{\beta}'_i \mathbf{x}, \quad i = 1, \dots, k \quad (13)$$

where

$\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_k$ are the k vector of intercepts for the

4 Factors

$\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_k$ are the k vectors of the slope

parameters

x_1 =Everyday Math

x_2 =Investigations

x_3 =Trailblazers

$$x_4 = \text{HM Math}$$

Results from Model (2) indicate the relative “preference” for students using a given textbook to choose distractors associated with each factor.

The results from Models (1) and (2) show how the each textbook impacts the odds of CFG membership for the distractors chosen, followed by how each textbook is related to the choices of distractors with respect to the factors representing the item attributes.

CHAPTER 4 : RESULTS

The analysis described in Chapter 3 indicates that CFG membership is not likely to occur by chance alone and that the TFGs replicate the CFGs and are consistent across textbooks. The focus of the analysis turns toward the impact of item attributes and textbook exposure on CFG membership. Using data collected from teacher consultations in combination with district, school, student, item, and distractor data, this chapter explores the relationship between these variables and CFG membership.

Phase I – Common Foil Groups

Distractors in CFG 1 come from items covering each of the five content strands listed in Table 2.1; however, the mistakes represented by these distractors are sometimes related despite the differences in what the items are measuring. The most common mistakes in CFG 1 are not computational in nature; they relate to definitions, terminology, and the ability to identify the appropriate operation or formula to solve the problem. The teacher experts agreed, for example, that the items related to fractions and computation of perimeter and area related to students not knowing definitions or formulas rather than the students making arithmetic errors while using correct formulae.

Distractors relating to fractions involved confusing the numerator and denominator, choosing the reciprocal of the correct answer, or choosing the incorrect operation (+ over -) from a figure. Similarly, when asked to read scales on axes and to interpret graphs, distractors in CFG 1 most often related to students confusing the max and the min, the max with the range, or confusing the horizontal with the vertical axis.

When students were asked to solve problems related to area and perimeter, students were likely to confuse the formula for area and perimeter or to use an incorrect

formula that was very close to the correct formula (adding $l+w$ rather than $2(l+w)$ to obtain the perimeter or choosing $l+w$ rather than $l \times w$ for area). Distractors in CFG 1 also included confusing the definitions of prism and pyramid in one item, and the definitions of line and line segment in another. Students also had trouble identifying the freezing and boiling points of water in degrees Fahrenheit.

The mistakes represented by the distractors making up CFG 1 are most related to identification and recognition rather than computation, and occur consistently across mathematical content. While it may seem that students choosing distractors from CFG 1 have difficulty mostly with fractions and perhaps geometry and measurement, further inspection of the specific distractors these students were most likely to choose reveals the link between the content of the incorrect items; this link specifically points to difficulty with definitions and appropriately choosing operations and/or formulae.

The remaining CFGs contain three or four distractors per group and do not contain multiple distractors from the same item. Distractors in the remaining groups come from items with generally higher p-values than those in CFG 1 and are chosen less frequently by higher achieving students and in smaller proportions than those in CFG 1. Some of the distractors in the remaining groups contain similar mistakes in recognition and identification to those in CFG 1; however, these groups contain more computationally based mistakes, specifically with time and money, as well as standard units of measure and number sense. In order to fully understand the differences between CFG 1 and the remaining groups, as well as the differences among the remaining groups, it is helpful to examine them with respect to different curricula.

Phase II – Textbook Foil Groups

Results based on the analyses described in Chapter 3 indicate that we may reject the null hypothesis that subgroup processes are equal to zero for the districts using a common textbook, but not for the districts using mixed texts. The distractors chosen and number of students per TFG are consistent within schools using the same textbooks and somewhat different between those using alternative texts.

The CFGs are based on the co-occurrences of distractors across the entire sample of students. The TFGs are based on the co-occurrences of students and distractors and are identified on a textbook-by-textbook basis. The results of both analyses yielded similar results; all results beyond this section will deal exclusively with the CFGs.

TFGs identified during Phase II of the analysis closely resemble the CFGs identified during Phase I. For each textbook there was one large group of common foils, followed by several smaller foil groupings. The large foil groupings for all of the textbooks resemble CFG 1 with minor variations. The smaller TFGs tend to resemble the remaining 12 CFGs; deviations from the CFGs represent the effect of textbook exposure on the distractor choices.

Correspondence Analysis

Correspondence analysis is a descriptive technique that allows for analysis of two-way tables representing counts of CFGs and textbooks. Results from the CA provide information on how the CFGs relate to each textbook, and provide a visual representation of how “close” different CFGs are to textbooks. For the first CA, the data are arranged with columns representing the four textbooks and rows representing the 13 CFGs. Table 4.1

provides the cell frequencies for the two-way table representing CFGs (rows) by Textbooks (columns).

The goal of CA is to find lower dimensional representations of the original contingency table that still retain all, or almost all, of the information about the differences between CFGs and textbooks.

Table 4.2 provides: the singular values; Eigen values; percentages of inertia explained; cumulative percentages; and the contribution to the overall Chi-square. Three dimensions account for 100% of the overall Chi-square distance.

Table 4.1: Two way Contingency Table for CFG by Textbook

CFG/Textbook	EM	INV	TB	HM	Total
1	1045	3360	4563	1588	10376
3	46	196	299	109	650
4	46	105	167	53	368
10	64	193	307	78	642
11	31	132	174	47	384
13	84	255	243	122	704
14	54	198	276	61	589
15	45	123	201	70	439
16	78	236	292	102	708
17	37	173	193	70	473
18	81	283	314	153	831
19	69	237	316	92	714
38	69	237	316	92	714
Total	1722	5667	7320	2602	17311

Table 4.2: Eigenvalues and Inertial for all Dimensions

Input Table (Rows x Columns): 13 x 4

Chi-square=85.99 Degrees of Freedom=36. $p>0.001$

No. of Dimensions	Singular Value	Principal Inertia	Chi-Square	Percent	Cumulative Percent
1	0.055	0.003	52.64	61.21	61.21
2	0.034	0.001	19.84	23.08	84.29
3	0.028	0.0008	13.51	15.71	100.00
Total		0.005	85.99	100	

Since the dimensions are extracted to maximize the distances between the row and column points, each successive dimension will explain less and less of the overall Chi-square value. Three dimensions show which textbooks are “closest” to each CFG. The results of the CA are provided in Table 4.3 and indicate INV and HM are on the same dimension, while EM and TB represent the remaining dimensions.

Table 4.3: Correspondence Analysis Results

CFG by Text Variable	Value	Dimension
CFG	10	1
CFG	13	1
CFG	14	1
CFG	18	1
Text	TB	1
CFG	1	2
CFG	3	2
CFG	15	2
CFG	19	2
CFG	38	2
Text	HM	2
Text	INV	2
CFG	4	3
CFG	11	3
CFG	16	3
CFG	17	3
Text	EM	3

Results from Table 4.3 show how each CFG is most related to each textbook, keeping in mind that the combined responses of all students from each specific textbook grouping showed that nearly every single distractor was chosen. These results show which CFGs and texts are most related based on expected and observed cell counts. These results also correspond to the results of the TFG analysis with respect to the types of mistakes students were likely to make based on his or her textbook.

Dimension 1 includes CFGs 10, 13, 14, and 18 and is related to the TB textbook. Mistakes common to these CFGs include recognizing the correct operation, specifically choosing addition or subtraction over multiplication or division, respectively; comparing and ordering numbers and calculating perimeter.

Dimension 2 includes CFGs 1, 3, 15, 19, and 38 and is related to both HM and INV textbooks. In addition to the most common mistakes (CFG 1), students using these

textbooks and located in these CFGs tend to have mistakes related to identifying the correct operation (adding, subtracting, multiplying and dividing), adding and subtracting money in mixed units, and understanding common units of measurement.

Lastly, Dimension 3 contains the EM textbook and CFGs 4, 11, 16 and 17.

Distractors in this dimension represent mistakes mostly related to identifying and classifying 2-D and 3-D shapes and adding and subtracting money in dollars and cents.

Figure 4.1: Correspondence Analysis of CFG by Textbook – Dimension 1 vs.

Dimension 2 provides the locations of CFGs and textbooks on each of the three dimensions.

Textbooks are indicated by the (+) symbol while CFGs are labeled with the (○) symbol.

These figures help to see which textbooks and CFGs are spatially close to one another.

Some CFGs and textbooks located on the same dimension will have opposing signs on the coordinate values. This shows the variability within dimensions and across dimensions.

Figure 4.1 shows the CFGs and textbooks for Dimension 1 (D1) versus Dimension 2 (D2). D1 contains the TB textbook and CFGs 10, 13, 14, and 18. CFGs 13 and 18 have positive coordinates on D1, while CFGs 10 and 14 and the TB textbook have negative coordinates. Distractors in CFGs 10 and 14 are alternative distractors to the same items' distractors in CFGs 13 and 18. Students using the TB textbook were likely to miss the items in these CFGs, and are more likely to make the mistakes represented in CFGs 10 and 14 rather than in 13 and 18.

Common items for all four of the CFGs in D1 include: recognizing multiplication and division situations; composing/decomposing triangles and rectangles; comparing/ordering numbers up to 10,000; adding/subtracting money in dollars and cents; and calculating the

perimeter/area of squares and rectangles. The distractors chosen between these CFGs are not the same; this is indicated by the distance between these CFGs along D1.

Students in CFGs 10 and 14 were likely to make consistent mistakes on items of similar content. For the item asking students to order three numbers, they consistently chose numbers outside of the range, and above the largest number. Conversely, students in CFGs 13 and 18 were likely to choose numbers outside the range and below the smallest number. Students in all of the CFGs on D1 had problems identifying the proper formula for computing perimeter for squares and rectangles. Students choosing distractors in CFG 13 found the perimeter of a square with sides measuring five inches to be 30 inches, while students in CFG 14 added the length and width, but forgot to multiply this value by two.

D2, also shown in Figure 4.1, contains CFGs 1, 3, 5, 19, and 38, as well as the INV and HM textbooks. CFGs 3 and 15 and HM have positive D2 coordinates and are the most closely linked within D2. CFGs 38 and 19 are closest to the INV textbook and have negative coordinates on D2. CFG 1 is also included here, but is quite close to zero (and on all other dimensions as well, since it represents the mistakes common across all students).

The mistakes contained in CFGs 38 and 19, most closely associated with the INV textbook, include: using relationships between sizes and standard units; using common measures of length, time, and weight; knowing benchmark temperatures; and distinguishing between units of length and area.

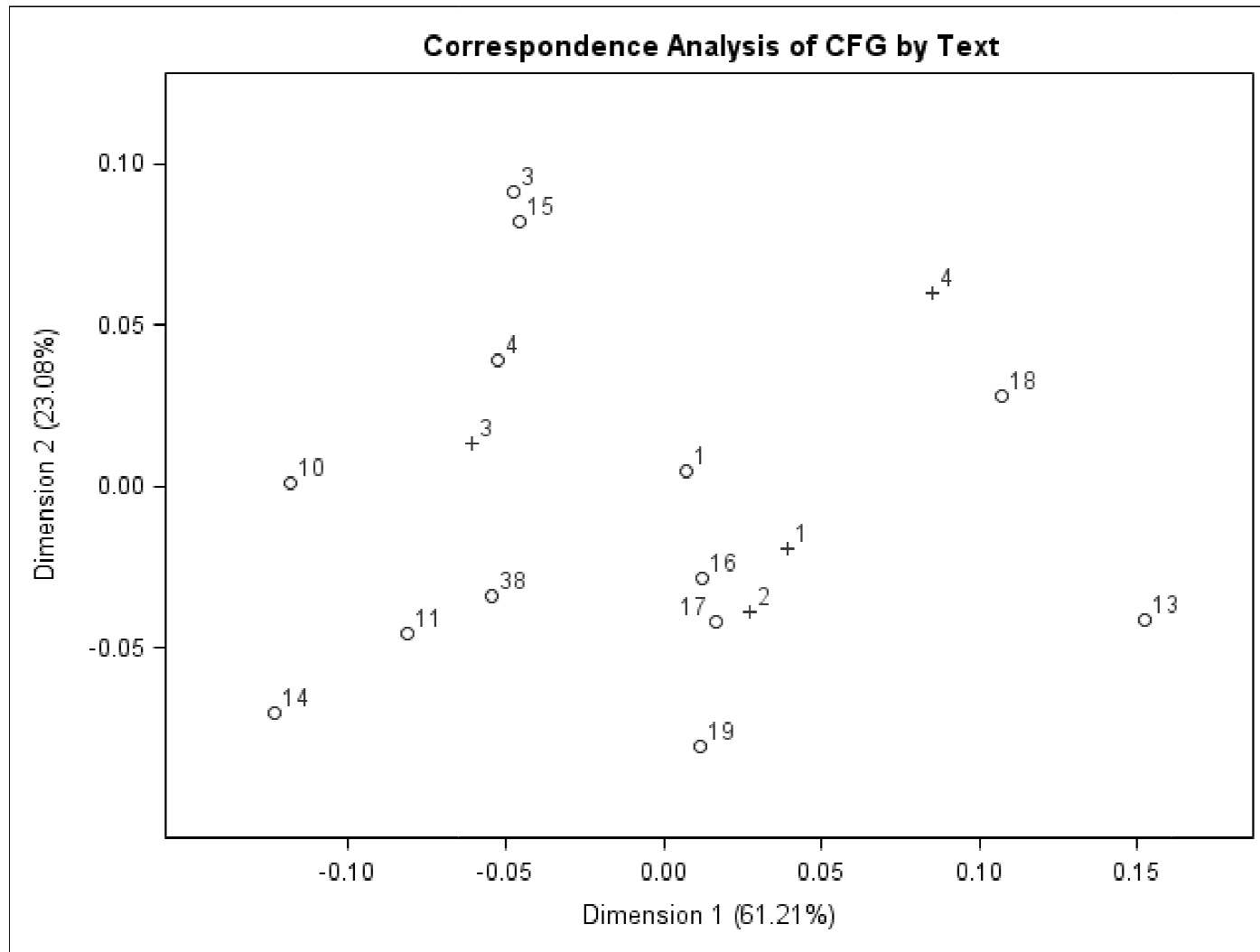
The mistakes in CFGs 3 and 15, and most closely related to the HM textbook include identifying the operation, solving the problem, and understanding the meaning of decimals related to money. The nature of the mistakes suggests students have a difficult time deciding the proper operation and carrying it out.

The third dimension, D3, contains the EM textbook and CFGs 4, 11, 16, and 17. EM and CFGs 4 and 16 have positive D3 coordinates while CFGs 11 and 17 have negative D3 coordinates. D3 represents mistakes involving identifying, classifying, describing, and comparing familiar 2-D and 3-D shapes, adding and subtracting money in dollars and cents, and understanding the meaning of 0.50 and 0.25 related to currency.

CFGs 4 and 16 indicate misconceptions around modeling fractions; both on the number line and with respect to adding and subtracting money and adding with mixed units. CFGs 11 and 17 contain similar content, and also include mistakes related to reading charts and adding/subtracting with regrouping. Item 35, which covers understanding fractions related to money, has distractors represented in CFG 4 and CFG 11. We can see why these groups are on the same dimension, but have different values for their respective coordinates.

The CA provides an illustration of the variations seen in the TFGs between textbooks by showing each textbook's "proximity" to the CFGs. The CFGs and textbooks associated with each dimension represent the variation in TFGs at the district levels and help to link the results of the CFGs to TFGs and to textbooks. The CA helps to show that the TFGs, based on distractors of choice, may be similarly represented by CFGs, which are based upon the co-occurrences of distractors, when controlling for textbook use.

Figure 4.1: Correspondence Analysis of CFG by Textbook – Dimension 1 vs. Dimension 2



Here (+1) represents EM, (+2) Investigations, (+3) Trailblazers, and (+4) represents TB.

Phase III – Principal Factor Analysis and BCLM Logistic Regression

Teacher Consultations

The following discussions for each Textbook are synthesized from conversations with all of my teacher experts (Phone conversations, September 2008-2010). The largest concerns, regardless of the textbook are: (1) the spiraled curriculum disadvantages lower SES students as they are more likely to move and receive an even further fragmented mathematics learning experience, (2) the time and resources necessary to prepare for daily coursework and therefore to comply fully with each curriculum's design is overly burdensome, and (3) the various ways students may learn mathematics may not be accurately represented on a standardized assessment.

Textbook and CFG expectations

Everyday Mathematics presents the most controversial approach to teaching mathematics (according to the teacher experts), partially because of its use of a distributed practice model. This model spreads content out over several lessons, to be introduced and then re-introduced later. The authors site research in support of their methods, as noted in the analysis section. When interviewing teachers regarding the efficacy of the EM curriculum, teachers either really liked the program, or were not comfortable with it at all; no teachers felt undecided about it. Most of the teacher reactions suggest that EM is wonderful for some students, and very difficult for others. For example, Ms. E suggested "I tutor students for the State test in Illinois, and some students who would be lost learning traditional algorithms excel with the Everyday Math curriculum, but there are some kids who move around a lot and end up in different schools studying different units, and how do I explain this out of the blue?" Ms J, also an Illinois teacher stated, "When I first learned

about Everyday Math, I was not confident in my abilities to teach these new methods, or how kids would react. I was really against the lattice multiplication, but the more I teach it, the more I like it.”

The largest complaint from teachers relates to the amounts of preparation, the lack of focus on standard units of measurement (students use block-based manipulatives), and the use of lattice multiplication. At the same time, teachers supporting EM tout these attributes as the strengths of the curriculum. In the latter editions of EM, additional algorithms for computing addition and multiplication were added to the text at the request of teachers unsure about the newer methods. Mrs. M, from Michigan notes, “When they [the publishers] send you the totes full of materials for each lesson, the stacks and stacks of totes are overwhelming. I do my best to try to get most of the materials ready for each lesson, but I have to teach all other subjects too, and sometimes, and I’m not alone, I don’t do all of the preparation or manipulative activities I’m supposed to.”

Math Trailblazers, on the other hand, sites the TIMSS study directly in the explanation of their philosophical basis for the text. TB aims to provide strong conceptual foundations and skill development through examining mathematics and science together, but in a way consistent with the National Standards. Teachers reported liking the TB textbooks, but only after revisions providing more instructional support. The main complaint about the use of TB came from parents frustrated by the lack of emphasis put on traditional problem solving. They bemoaned the fact that TB directs students to discuss how, theoretically, one would go about solving a given problem. Teachers were also skeptical of allowing students to invent their own algorithms without earlier intervention than the TB curriculum suggests. Finally, teachers voiced concerns about the use of scoring

rubrics rather than traditional assessments, if for no other reason than the MEAP is not scored using rubrics.

The *Investigations* textbooks are almost a hybrid between EM and TB. The INV curriculum, like the TB curriculum encourages students to invent their own algorithms to complete problems. Unlike TB, students are expected to actually solve the problems and not simply discuss possible solutions. Similar to EM, topics are spiraled, or introduced in waves, with the goal of creating meaningful connections between different mathematical content areas.

The *Harcourt Math 2004* textbooks do reflect the National standards, but present mathematics in a more traditional method than the NSF-reform texts. Although the HM texts do focus on concepts and algorithms, they are designed with several interactive labs, and with a focus on combining new and old methods, letting students choose the most suitable strategy.

Each of the teachers had expectations for the types of mistakes students would make, based on the textbook he or she used in class. During a conference call with all of the teacher experts, I asked them to guess what types of item attributes were likely to be associated with different types of math textbooks. The teachers agreed that students using EM, INV, and TB were likely to do well on items involving place value, number sense, arithmetic, and word problems involving time and money. These students were likely to have trouble with items requiring recall of formulae and items requiring multiple steps to solve problems, this being especially true for the rubric based INV. Items requiring students to read and use charts, add and subtract fractions with different denominators, and calculate area or volume, teachers agreed, would be difficult for all students at this age,

regardless of textbook. None of the teachers specifically mentioned how they believed textbooks would affect students' ability to recognize and/or manipulate two-dimensional or three-dimensional shapes.

Prior to examining the results of the PFA, I reviewed the results of the CA to see if the teacher expectations aligned with the results. As expected, EM and TB are located on separate dimensions, and INV is located on the third, albeit with HM. Teachers thought EM students would not have trouble with items involving place value and money, while the CA reveals that students did have issues with these items. It is important to note that outside of CFG 1, the distractors listed are more frequently chosen by students in the lower 33% of the score distribution. Teachers agreed that, for all textbooks, the CFGs make sense for the middle and lower performing students along the distribution of total scores.

Teacher Reactions to Correspondence Analysis

Expectations for the likely mistakes for TB students were very close to the observed CA results. TB students had difficulty related to choosing the correct operation for arithmetic, comparing numbers, and composing/decomposing geometric figures. Teachers are quite critical of the rubric-based curriculum which asks students to create essays rather than solve problems, and were not surprised by the results.

Teachers were surprised by the similarity or loading onto the same dimension for the INV and HM textbooks; teachers were also surprised by the difficulty students using these texts had with adding and subtracting dollars and cents. Teachers weren't surprised with the tendency for students choosing distractors related to CFGs in Dimension 3 related to identifying differences between common units of length or identifying the correct operation to solve problems. Teachers expected this for INV students, based on the focus of

students inventing their own algorithms. Teachers were surprised that HM students seemed to make similar mistakes since HM incorporates exploratory and rote methods of teaching and learning.

The teacher expectations regarding patterns of mistakes were very close to the patterns revealed during the analysis. The distractors chosen and mistakes made by textbook aligned with the teacher expectations, and the salient Process and Skill Factors identified by the PFA also corroborate the teacher expectations. The factors associated with each CFG seemed to have more agreement with the teacher expectations; however, the CA results support expectations for the lower performing students, while the PFA results tend to reflect expectations for the entire distribution of student achievement.

Factor Analysis

Using the counts of attribute codes associated with each distractor chosen by each student, the PFA retained four factors. The results of the PFA using a Promax, oblique rotation are displayed in the tables and figures below. Table 4.4 provides the eigen values of the reduced correlation matrix, and it shows that Factor 1 accounts for almost 51% of the variance, Factor 2 accounts for 23%, Factor 3 accounts for 19% and Factor 4 accounts for 11%.

**Table 4.4: Eigenvalues of the Reduced Correlation Matrix:
Total=6.06 Average=0.55**

Factor	Eigenvalue	Difference	Proportion	Cumulative
1	3.11	1.70	0.51	0.51
2	1.42	0.29	0.23	0.75
3	1.13	0.47	0.19	0.93
4	0.66	0.30	0.11	1.04
5	0.36	0.23	0.06	1.10
6	0.12	0.09	0.02	1.12
7	0.04	0.15	0.01	1.13
8	-0.11	0.05	-0.02	1.11
9	-0.16	0.07	-0.03	1.08
10	-0.23	0.05	-0.04	1.05
11	-0.28		-0.05	1

Figure 4.2: Scree and Variance Explained Plots shows the Scree and Variance Explained plots. The plots indicate that the first four factors with eigen values greater than one account for over 100% of the variance (104.1%). This is possible since the reduced correlation matrix is not positive definite, thus the negative eigen values. By default, PROC FACTOR retains four factors; the first factors accounting for at least 100% of the variance are retained.

Figure 4.2: Scree and Variance Explained Plots

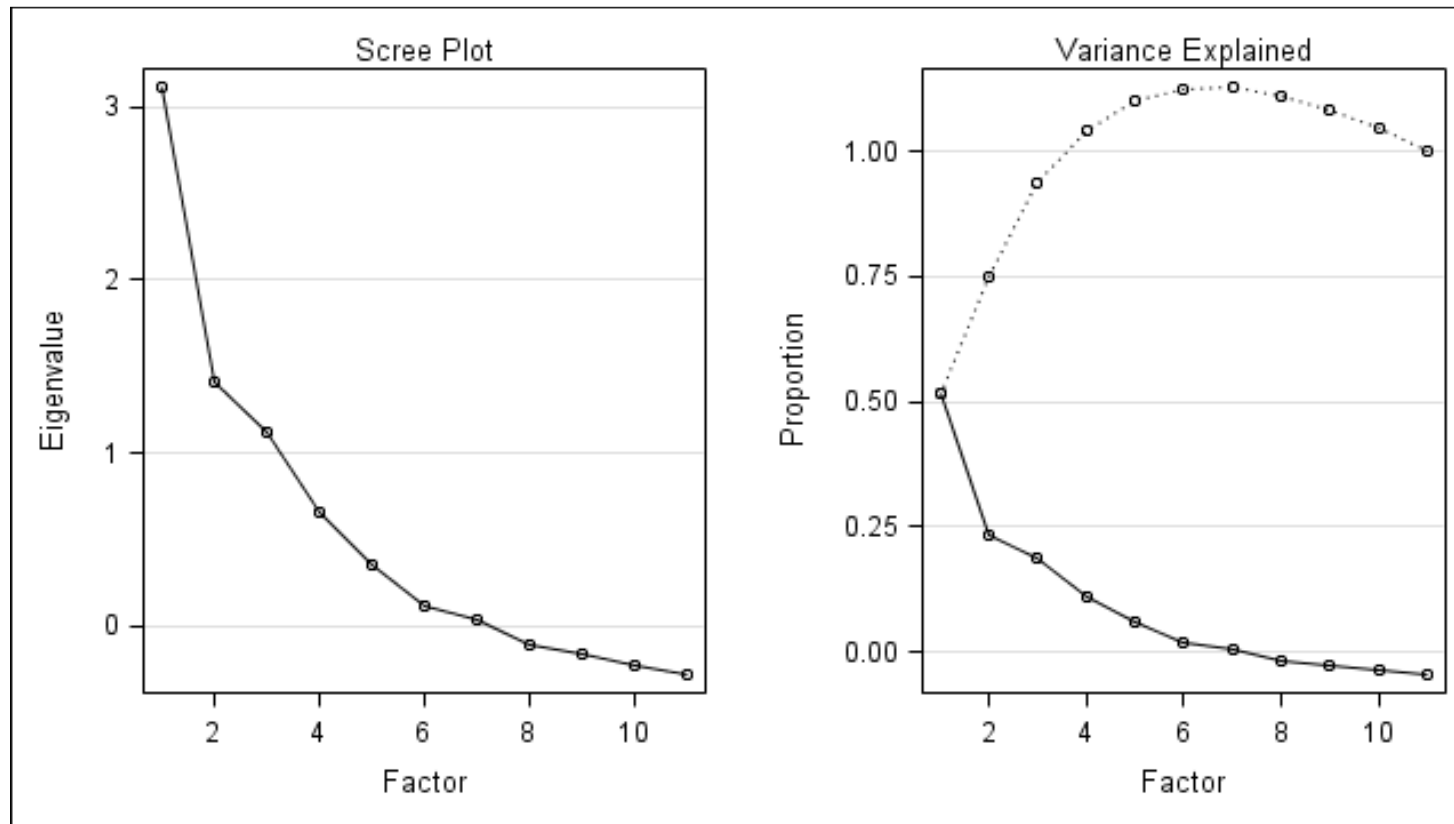


Table 4.5 provides the final rotated solution for the PFA. Factor 1 includes two process and two skill attributes: (P7) Generating, visualizing, and reading figures in graphs; (P9) Management of data and procedures; (S3) Using figures, tables, charts and graphs; and (S9) Comparison of two or more entities. Factor 2 includes two process attributes and no skill attributes: (P1) Translate/formulate equations and expressions to solve a problem, and (P2) Computational applications of knowledge in arithmetic and geometry. Factor 3 contains one process and one skill attribute: (P8) Applying and evaluating mathematical correctness; and (S2) Apply number properties and relationships; number sense/number line. Factor 4 has two skill attributes and one (negative) process attribute: (S5) Evaluate/verify/check options; (S8) Solving novel or unfamiliar problems and (-P10) (the absence of items containing this code) Quantitative and logical reasoning.

These factors represent the underlying nature of the distractors chosen by each of the students. The item attributes associated with each item provide a more meaningful understanding of the nature of the patterns of mistakes students made. The attractiveness for one CFG over another for students using different textbooks may be more clearly understood by examining the types of processes and skills associated with each distractor.

Table 4.5: Rotated Factor Structure

Attribute	Factor1	Factor2	Factor3	Factor4
s3	79	7	-11	9
p7	69	7	-21	30
p9	65	-3	2	-36
s9	59	-9	24	-18
p2	-5	86	-12	-1
p1	7	81	11	-4
p8	4	-3	79	17
s2	0	2	70	-3
s5	5	12	11	62
s8	1	-36	-9	62
p10	31	-19	-14	-52

Using the results from the PFA, the final analysis models the odds of CFG membership based on textbook use followed by the odds of a student associating with each factor based on textbook use. The combined results show the relationships between CFGs and textbooks and how textbooks are related to each of the factors. These models provide a comprehensive understanding of how CFGs are related to textbooks and how certain combinations of item attributes are associated with textbooks.

BCLM Logistic Regression

Model (1) CFGs and Textbooks

Recall from Equation (12) that Model (1) examines the odds of CFG classification based on the textbook a student used. The reference CFG is CFG 1 and the reference textbook is “mixed use”. Table 4.6 provides the model fit statistics and Table 4.7 provides the results of the hypothesis test determining whether or not Beta is significantly different from zero.

Table 4.6: Model (1) Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	74394	74365
SC	74491	74846
-2 Log L	74370	74245

Table 4.7: Model (1) Null Hypothesis, Beta=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood			
Ratio	125	48	<.0001
Score	122	48	<.0001
Wald	121	48	<.0001

These results indicate that the model fits the data well (larger values indicate good model fit) and that we may reject the null hypothesis of Beta being equal to zero.

Textbooks are significant predictors of CFG membership, and the results in Table 4.8 show odds ratios for each textbook and CFG where that odds ratio is significant at a 0.15 level or smaller. To interpret the results in Table 4.8, it is important to remember that CFG 1 and the category for mixed use textbooks are the respective reference categories. All of the intercepts for each CFG are statistically different from zero, but not all textbooks had unique effects on CFG membership. Results shown in Table 4.8, therefore, do not include odds-ratio estimates for textbook effects where none are significant.

The point estimates indicate the increase in the odds for CFG based on textbook use. Students using EM were more likely to choose distractors from CFG 13 and less likely than students in the mixed group to select distractors from CFGs 11 and 1. Students using the INV textbook were less likely than students in the mixed group to select distractors related

to CFGs 3, 10, 14, and 15. HM students were more likely to choose distractors from CFGs 1, 13, and 18 but less likely to choose distractors from CFG 14. Finally, the results from this analysis suggest that TB students are no more likely to select distractors from any CFG, relative to students using mixed textbooks.

The results from Model (1) show how each textbook affects the odds of CFG membership. Model (2) examines the relationship between the factors with which the distractor choices are associated and the textbook a student uses. The reference categories for Model (2) are the categories representing mixed textbook use and all distractors not related to the four defined factors.

Table 4.9 provides the model fit statistics and Table 4.10 provides the results of the hypothesis test determining whether or not Beta is significantly different from zero. These results indicate that the model fits the data well and that the null hypothesis that $\text{Beta}=0$ may be rejected. The textbook a student uses not only affects the preference for students to choose distractors within certain CFGs, but also affects the preference for students to make mistakes associated with the four factors identified in the PFA.

Table 4.8: Odds Ratios for CFG Membership by Textbook with Wald Confidence Intervals

Parameter	CFG	DF	Estimate	Error	Chi-Square	Pr > ChiSq	Exp (Estimate)	95% Confidence Limits	
EM	3	1	-0.4383	0.2135	4.2118	0.0401	0.645	0.425	0.98
EM	11	1	-0.3509	0.2428	2.0873	0.1485	0.704	0.437	1.133
EM	13	1	0.276	0.1906	2.0971	0.1476	1.318	0.907	1.914
INV	3	1	-0.2227	0.1367	2.6542	0.1033	0.8	0.612	1.046
INV	10	1	-0.2699	0.1365	3.911	0.048	0.763	0.584	0.998
INV	14	1	-0.2172	0.1365	2.5333	0.1115	0.805	0.616	1.052
INV	15	1	-0.2432	0.1559	2.4343	0.1187	0.784	0.578	1.064
HM	1	1	0.2653	0.1289	4.2361	0.0396	1.304	1.013	1.679
HM	13	1	0.3615	0.1702	4.5085	0.0337	1.435	1.028	2.004
HM	14	1	-0.4483	0.1913	5.4892	0.0191	0.639	0.439	0.929
HM	18	1	0.4621	0.1634	7.9948	0.0047	1.587	1.152	2.187

Table 4.9: Model (2) Fit Statistics

Criterion	Only	Covariates
AIC	69334	69248
SC	69374	69448
-2 Log L	69324	69198

Table 4.10: Model (2) Null Hypothesis, Beta=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood			
Ratio	126	20	<.0001
Score	117	20	<.0001
Wald	113	20	<.0001

Table 4.11: Odds Ratios for Factor Association by Textbook with Wald Confidence Intervals

Parameter	Factor	DF	Estimate	Error	Chi-Square	Pr > ChiSq	Exp(Est)	Point Estimate	95% Confidence Interval	
EM	1	1	-0.6901	0.1212	32.418	<.0001	0.502	0.502	0.395	0.636
INV	1	1	0.2048	0.0665	9.4813	0.0021	1.227	1.227	1.077	1.398
INV	2	1	-0.1387	0.0488	8.0751	0.0045	0.87	0.87	0.791	0.958
INV	4	1	-0.1555	0.0848	3.362	0.0667	0.856	0.856	0.725	1.011
INV	2*3	1	-0.14	0.0699	4.005	0.0454	0.869	0.869	0.758	0.997
TB	2	1	-0.1506	0.0461	10.6858	0.0011	0.86	0.86	0.786	0.941
TB	4	1	-0.1486	0.0797	3.4764	0.0623	0.862	0.862	0.737	1.008
HM	1	1	0.2132	0.0819	6.771	0.0093	1.238	1.238	1.054	1.453
HM	2	1	-0.0973	0.0612	2.5281	0.1118	0.907	0.907	0.805	1.023

The results of the network analysis identify which distractors are likely to be chosen together, in general and with respect to the textbook a student uses in the classroom. The Correspondence Analysis showed which textbooks were dimensionally closest to which CFGs and how the CFGs chosen in common for a given textbook relate to one another. The Logistic Regression analysis shows the effects of textbooks on the CFGs and on the Factors representing the different groupings of item attributes. The following section combines the results from the CA and both models in the logistic regression.

Everyday Math

Results found in Table 4.3 suggest that the EM textbook shares a dimension with CFGs 4, 11, 16, and 17. Results from Model (1) in Table 4.8 indicate further that EM students are less likely when compared to the control group to select items from CFGs 3 and 11 and are more likely to choose items from CFG 13. In Model (2) students using the EM textbook are less likely to select items from Factor 1 than the control group. Since Factor 1 is only associated with CFGs 1 and 13, the analysis for the EM textbook suggests that students are less likely to choose items from CFG 1 than CFG 13, with respect to Factor 1 and are most associated with CFGs 4, 11, 16 and 17.

Investigations

The CA results provided in Table 4.3 identified CFGs 1, 3, 15, 19, and 38 in close proximity to the INV textbook. Parameter estimates from the logistic regression suggest somewhat different results, as Model (1) results in Table 4.8 suggests INV students are less likely than students using Mixed Texts to choose distractors in CFGs 3, 10, 14, and 15. The logistic regression for Model (2) estimates provided in Table 4.11 suggests that the odds

for an INV textbook user are 1.2 times larger than that of students using mixed textbooks for selecting distractors from Factor 1. Students using the INV textbook are most likely to select distractors from CFGs 1, 3, 19, and 38 while having the largest propensity to select items related to Factor 1.

Harcourt Math

Students using the HM textbook were identified along with INV students in the CA provided in Table 4.3. Parameter estimates from Model (1) in Table 4.8 suggest HM students are more likely than mixed-use students to select distractors from CFGs 1, 13, and 18. HM students are less likely than the control group to select distractors from CFG 14. Results from Model (2) in Table 4.11 indicate that HM students are more likely than the reference group to select distractors related to Factor 1 and less likely to select items from Factor 2. HM students were most likely to select items from CFGs 5, 15, 19, and 38.

Trailblazers

The CA and suggests that TB students are most likely to choose distractors related to CFGs 10, 13, 14, and 18 as shown in Table 4.3. Results from Model (1) in Table 4.8, reveal that the logistic analysis does not suggest increased odd for TB students to prefer distractors located within any one CFG over another. The results from Model (2) in Table 4.11 suggest that TB students were less likely than students using mixed textbooks to select distractors related to factors 2 and 4.

CHAPTER 5 : DISCUSSION

The goal of this dissertation was to determine whether statistically significant groupings of distractors and students existed, based upon responses to a large-scale multiple choice mathematics assessment. Using the network analysis, correspondence analysis, and logistic regression, a significant relationship was shown between the types of mistakes students are likely to make and the textbook used in their classroom. Recall the hypothesis:

Hypothesis I: CFGs and TFGs do not occur randomly and are meaningful with respect to known cognitive correlates of mathematical mistake making.

Hypothesis II: TFGs should be consistent among schools using the same textbook and different for schools using other textbooks, while maintaining an overall relationship to the CFGs.

Conclusions and Discussions

Examining the z-scores from the likelihood ratios computed for each school, I verified Hypothesis I, with the caveat that the algorithm does not identify CFGs for schools reporting mixed textbook use. Additionally, TFG positions containing less than four students included students in the upper or lower tails of the score distributions, by school. That is, students scoring above 95% correct and below 50%. These CFG positions were omitted from the analysis.

Simultaneously examining the CFGs and the TFG positions shows the mistakes students across districts and schools are likely to make, while enabling the comparison of schools using different textbooks. CFG 1 contains the distractors students in the sample were most likely to choose, regardless of curricula. The variation in CFG combinations

making up each individual TFG reflect the differences in items and distractors occurring by textbook use.

Using existing literature and input from teacher experts, the differences in CFGs between schools using different textbooks highlight both the strengths of each representative textbook, as well as the relative weaknesses.

In order to solve a mathematics problem students need to be able to simultaneously recognize the symbols and/or natural language presented in an item's stem choose the correct operation, and then correctly carrying out that operation or procedure (Duval 2006). Duval explains the cognitive paradox of accessing and applying, simultaneously, knowledge object in mathematics: (1) in order to do any mathematical activity, semiotic representations must necessarily be used even if there is the choice of the kind of semiotic representation, but (2) the mathematical objects must never be confused with the semiotic representations that are used (2006).

One of the key differences in each of the NSF reform texts as well as the HM text is the way these representations are presented to students, and how students are expected to learn and understand conceptual and algorithmic strategies. The presentation of mathematics content varies widely from EM, using (*) rather than ($\times$) for multiplication, as well as presenting multi-digit multiplication in lattice form, rather than in traditional column form. For multiplication, specifically, the mistakes made by the place-value-oriented approach will be different from those made by students using the traditional column methods (Fuson and Briars 2000). These strategies will also be different from those learned, for example, in the TB curriculum, as students are taught to group things in hundreds and tens, then to tack on the ones.

Given the various different approaches to presenting the material and the different types of manipulatives and computational strategies, it is difficult to imagine the demands on test designers to create items and distractors appropriate for each type of strategy. If, however, it is possible to examine the problems giving students trouble in each school, combined with the specific distractor chosen, it is possible to design assessments to take these curricular differences into account.

When examining the second hypothesis, we can see that the type of textbook affects how students respond to test items, and that the types of mistakes made reflect the foci of each respective textbook. While it is not clear that one textbook is superior to another, it is important to recall that each of the districts in this sample performed very well overall and compared to the rest of the students in the State.

The combined results of the CA and logistic regression analysis shows how textbook are likely to affect the selection of distractors and how those distractors relate to the underlying nature of the mistakes based on the association of each item with one of the four factors identified in the PFA. The combined results across the analyses show that there are meaningful groupings of distractors which are likely to be chosen together. These groupings are likely to be consistent across textbooks and to vary between textbooks. In addition to the differences in CFG selection by textbook, the factors or underlying nature of the mistakes are also likely to differ based on the textbook in question.

The way mathematics material is presented to students affects the way students approach solving problems on a State assessment. While the scores of the students using different textbooks may not be statistically different from one another, the nature of the mistakes likely does vary across texts, therefore giving a teacher the ability to correct

misconceptions. In order to ensure that data collected from such assessments are valid for each student, it is important for test makers to understand the different types of mistakes students can and do make based on the curriculum in the classroom. This ensures that the conclusions drawn and interventions designed for each student are appropriate.

Limitations

There are several realistic limitations to this research. First, due to the nature of the teacher and student data, I was not able to link direct teacher content coverage to specific students. In addition, using schools participating in the PROM/SE project limited my sample, but permitted me to examine the content coverage and textbooks used by each district. Although these districts are diverse, and the results are consistent between schools in a district, they are not a truly representative sample from the State. Further, students in these districts scored above the State averages, which may be due to the individual districts' commitment to mathematics and science improvement and/or the specific textbook in use. Additionally, this analysis examines only four textbooks on a restricted sample. It is possible that the types of mistakes students are likely to make are not fully represented by the textbooks examined in this model.

Despite these limitations, the analysis provides promising evidence that distractor analysis (with careful scrutiny regarding the content, process and skill codes, and textbook use, combined with social network models), can illuminate mistakes that students performing in the middle of the score distributions are likely to make. This, in turn, can provide teachers with information on the most frequently missed problems as well as the most frequent specific misconceptions, which may help inform future lesson plans.

Future Directions

In order to fully understand the differences between and the meaning of different KS positions, these analyses should be replicated using a larger sample with more complete textbook and content coverage information. These analyses would provide more evidence to suggest the strengths and weaknesses of each textbook, as well as solidifying the patterns of CFG positions for each. Additionally, analyses should be expanded to include schools using older texts and mixed tests to verify the lack of coherence in mistake making when a common curriculum is not used. The more we know about how students learning mathematics in different ways perform on standardized tests, the more information can be provided to test designers, item writers, and teachers in a classroom.

Future analyses might also consider examining the effects of the CFG positions by ethnicity and gender, as the goal of this dissertation was to examine the effects of curriculum on state assessments, using districts of varying levels of ethnic diversity and SES. Specifically, adding these variables into the analysis may provide additional insight into the way different types of students process different curricula. However, results in this dissertation suggest the effects are based on curriculum/textbook exposure rather than demographic categories.

REFERENCES

REFERENCES

- Baroody, A. J. (1984). Children's Difficulties in Subtraction: Some Causes and Questions. *Journal for Research in Mathematics Education*, 15(3), 203-213.
- Bell, M., & Bell, J. (Eds.). (1998-1996). *Everyday Mathematics* (1 ed.). Chicago: Wright Group McGraw-Hill.
- Ben-Zeev, T., & Star, J. R. (2001). Spurious Correlations in Mathematical Thinking. *Cognition and Instruction*, 19(3), 253-275.
- Carpenter, T. P., Fennema, E., & Peterson, P. L. (1989). Using knowledge of children's mathematics thinking in classroom teaching: an experimental study. [Feature]. *American Educational Research Journal*, 26, 499-531.
- Carpenter, T. P., Hiebert, J., & Moser, J. M. (1981). Problem Structure and First-Grade Children's Initial Solution Processes for Simple Addition and Subtraction Problems. *Journal for Research in Mathematics Education*, 12(1), 27-39.
- Carpenter, T. P., & Moser, J. M. (1984). The Acquisition of Addition and Subtraction Concepts in Grades One through Three. *Journal for Research in Mathematics Education*, 15(3), 179-202.
- Cobb, P., Wood, T. L., & Yackel, E. (1991). Assessment of a problem-centered second-grade mathematics project. [Feature]. *Journal for Research in Mathematics Education*, 22, 3-29.
- COMAP. (2003). The ARC Center Tri-State Student Achievement Study (pp. 59): National Research Foundation.
- Durans, N. J., & Holland, P. W. (1993). DIF Detection and Description: Mantel-Haenszel and Standardization. In P. W. Holland, H. Wainer & Educational Testing Service. (Eds.), *Differential item functioning* (pp. 35-66). Hillsdale: Lawrence Erlbaum Associates.
- Duval, R. (2006). A Cognitive Analysis of Problems of Comprehension in a Learning of Mathematics. *Educational Studies in Mathematics*, 61(1-2), 103-131.
- Fennema, E., Carpenter, T. P., & Franke, M. L. (1996). A longitudinal study of learning to use children's thinking in mathematics instruction. [Feature]. *Journal for Research in Mathematics Education*, 27, 403-434.

- Field, S., Frank, K. A., Schiller, K., Riegle-Crumb, C., & Muller, C. (2006). Identifying positions from affiliation networks: Preserving the duality of people and events. *Social Networks*, 28(2), 97-186.
- Frank, K. A. (1995). Identifying cohesive subgroups. *Social Networks*, 17(1), 27-56.
- Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P. M., Capizzi, A. M., . . . Fletcher, J. M. (2006). The Cognitive Correlates of Third-Grade Skill in Arithmetic, Algorithmic Computation, and Arithmetic Word Problems. *Journal of Educational Psychology*, 98(1), 29-43.
- Fuchs, L. S., Seethaler, P. M., Powell, S. R., Hamlett, C. L., & Fletcher, J. M. (2008). Effects of Preventative Tutoring on the Mathematical Problem Solving of Third-Grade Students with Math and Reading Difficulties. *Exceptional Children*, 74(2), 155-173.
- Fuson, K. C., & Briars, D. J. (2000). Using a base-ten blocks learning/teaching approach for first- and second-grade place-value and multidigit addition and subtraction. [Feature]. *Journal for Research in Mathematics Education*, 21, 180-206.
- Fuson, K. C., Wearne, D., Hiebert, J. C., Murray, H. G., Human, P. G., Olivier, A. I., . . . Fennema, E. (1997). Children's Conceptual Structures for Multidigit Numbers and Methods of Multidigit Addition and Subtraction. *Journal for Research in Mathematics Education*, 28(2), 130-162.
- Green, B. F., Crone, C. R., & Folk, V. G. (1989). A Method for Studying Differential Distractor Functioning. *Journal of Educational Measurement*, 26(2, The Test Item), 147-160.
- Hiebert, J., Carpenter, T. P., & Fennema, E. (1996). Problem solving as a basis for reform in curriculum and instruction: the case of mathematics. [Feature]. *Educational Researcher*, 25, 12-21.
- Hiebert, J., & Wearne, D. (1985). A Model of Students' Decimal Computation Procedures. *Cognition and Instruction*, 2(3/4), 175-205.
- Hiebert, J., & Wearne, D. (1993). Instructional tasks, classroom discourse, and students' learning in second-grade arithmetic. [Feature]. *American Educational Research Journal*, 30, 393-425.
- Hiebert, J., & Wearne, D. (1996). Instruction, understanding, and skill in multidigit addition and subtraction. [Feature]. *Cognition and Instruction*, 14(3), 251-283.
- Holland, P. W., & Leinhardt, S. (1981). An Exponential Family of Probability Distributions for Directed Graphs. *Journal of the American Statistical Association*, 76(373), 33-50.
- Houghton, Mifflin, & Harcourt. (2010). Harcourt Math 2004 Retrieved June 1, 2010, from <http://www.hmhschool.com/store/ProductCatalogController?cmd=Browse&subcm>

[d=LoadDetail&level1Code=3&ID=1007500000064247&sortEntriesBy=NAME&division=S01&gradeLevel=3&nextPage=School/browseby.jsp&frontOrBack=F](http://www.eric.ed.gov/fulltext/ED484841.pdf)

- Leinhardt, S., & Wasserman, S. S. (1979). Exploratory Data Analysis: An Introduction to Selected Methods. *Sociological Methodology*, 10, 311-365.
- McKnight, C. C., & Schmidt, W. H. (1998). Facing Facts in U.S. Science and Mathematics Education: Where We Stand, Where We Want To Go. *Journal of Science Education and Technology*, 7(1), 57-76.
- MSEB. (1990). Reshaping School Mathematics: A Philosophy and Framework for Curriculum. Corp Author(s): National Academy of Sciences - National Research Council, Washington, DC. Mathematical Sciences Education Board (pp. 73).
- NCTM. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, Va.: The Council.
- NCTM. (1991). Professional Standards for Teaching Mathematics. Corp Author(s): National Council of Teachers of Mathematics, Inc., Reston, VA (pp. 196). NCTM, 1906 Association Drive, Reston, VA 22091.
- NRC, & MSEB. (1989). Everybody Counts. A Report to the Nation on the Future of Mathematics Education. Corp Author(s): National Academy of Sciences - National Research Council, Washington, DC. Mathematical Sciences Education Board (pp. 131). Washington, D.C.: National Academy of Sciences
- National Research Council
- Mathematical Sciences Education Board.
- NRC, Shavelson, R. J., & Towne, L. (2002). *Scientific research in education*. Washington, DC: National Academy Press.
- Pettersson, A. (1991). Pupils' Mathematical Performance in Grades 3 and 6. A Longitudinal Study. *Educational Studies in Mathematics*, 22(5), 439-450.
- Powell, S. R., Fuchs, L. S., Fuchs, D., Cirino, P. T., & Fletcher, J. M. (2009). Effects of Fact Retrieval Tutoring on Third-Grade Students with Math Difficulties with and without Reading Difficulties. *Learning Disabilities Research & Practice*, 24(1), 1-11.
- PROM/SE. (2003) Retrieved June 2, 2010, from <http://promse.msu.edu>
- PROM/SE. (2006). Knowing mathematics: What we can learn from teachers. *PROM/SE Research Report Series, Vol. 2* Retrieved June 2, 2010, from http://www.promse.msu.edu/research_results/PROMSE_research_report.asp

- PROM/SE. (2009a). Content coverage and the role of instructional leadership. *PROM/SE Research Report Series, Vol. 7* Retrieved June 2, 2010, from http://www.promse.msu.edu/research_results/PROMSE_research_report.asp
- PROM/SE. (2009b). Opportunities to learn in PROM/SE classrooms: Teachers' reported coverage of mathematics content. *PROM/SE Research Report Series, Vol. 6* Retrieved June 2, 2010, from http://www.promse.msu.edu/research_results/PROMSE_research_report.asp
- PROM/SE. (2009c). Variation across districts in intended topic coverage: Mathematics. *PROM/SE Research Report Series, Vol. 5* Retrieved June 2, 2010, from http://www.promse.msu.edu/research_results/PROMSE_research_report.asp
- Resnick, L. B., Corp Author: Pittsburgh Univ, P. A. L. R., Development, C., & et al. (1989). Conceptual Bases of Arithmetic Errors: The Case of Decimal Fractions. Access ERIC: FullText. *Journal for Research in Mathematics Education, 20*(1), 8-27.
- Schmidt, W., Houang, R., & Cogan, L. (2004). A Coherent Curriculum: The Case of Mathematics. *Journal of Direct Instruction, 4*(1), 13-28.
- Schmidt, W. H. (1992). TIMSS Curriculum Analysis: Topic Trace Mapping. *Prospects, 22*(3), 326-333.
- Schmidt, W. H., & McKnight, C. C. (1998). What Can We Really Learn from TIMSS? *Science, 282*(5395), 1830-1831.
- Schmidt, W. H., Wang, H. C., & McKnight, C. C. (2005). Curriculum Coherence: An Examination of US Mathematics and Science Content Standards from an International Perspective. *Journal of Curriculum Studies, 37*(5), 525-559.
- Seethaler, P. M., & Fuchs, L. S. (2006). The Cognitive Correlates of Computational Estimation Skill among Third-Grade Students. *Learning Disabilities Research & Practice, 21*(4), 233-243.
- Tatsuoka, K. K., Corter, J. E., & Tatsuoka, C. (2004). Patterns of Diagnosed Mathematical Content and Process Skills in TIMSS-R across a Sample of 20 Countries. *American Educational Research Journal, 41*(4), 901-926.
- Thissen, D., & Steinberg, L. (1984). A response model for multiple choice items. *Psychometrika, 49*(4), 501-519.
- UCSMP. (2000). *Distributed Practice: The Research Base*. Paper presented at the Everyday Mathematics Leadership Institutes. http://everydaymath.uchicago.edu/about/research/distributed_practice.pdf
- UCSMP. (2010). Everyday Mathematics Website Retrieved June 1, 2010, from <http://everydaymath.uchicago.edu/>

- UIC. (2003). Research Foundations of Math Trailblazers Retrieved June 2, 2010, from <http://www.kendallhunt.com/uploads/2/MTBresearchbase.pdf>
- UIC (Ed.). (1997, 2004). *Math Trailblazers* (3rd ed.). Chicago: Kendall Hunt.
- van Galen, M. S., & Reitsma, P. (2010). Learning Basic Addition Facts from Choosing between Alternative Answers. *Learning and Instruction*, 20(1), 47-60.
- Wasserman, S., & Faust, K. (1994). *Social network analysis : methods and applications*. Cambridge ; New York: Cambridge University Press.