PREDICTING URBAN ELEMENTARY STUDENT SUCCESS AND PASSAGE ON
OHIO’S HIGH-STAKES ACHIEVEMENT MEASURES USING DIBELS ORAL
READING FLUENCY AND INFORMAL MATH CONCEPTS AND APPLICATIONS:
AN EXPLORATORY STUDY EMPLOYING HIERARCHICAL LINEAR MODELING

A dissertation submitted to the
Kent State University College and Graduate School
of Education, Health, and Human Services
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

By
Erich Robert Merkle
December 2010
A dissertation written by
Erich Robert Merkle

B.S., Heidelberg College, 1996
M.Ed., Kent State University, 1999
M.A., Kent State University, 1999
Ed.S., Kent State University, 2006
Ph.D., Kent State University, 2010

Approved by

______________________________
Caven S. Mcloughlin
Director, Doctoral Dissertation Committee

______________________________
Frank Sansosti
Member, Doctoral Dissertation Committee

______________________________
Jason D. Schenker
Member, Doctoral Dissertation Committee

Accepted by

______________________________
Mary Dellmann-Jenkins
Director, School of Lifespan Development and Educational Sciences

______________________________
Daniel F. Mahony
Dean, College and Graduate School of Education, Health and Human Services
Contemporary education is experiencing substantial reform across legislative, pedagogical, and assessment dimensions. The increase in school-based accountability systems has brought forth a culture where states, school districts, teachers, and individual students are required to demonstrate their efficacy towards improvement of the educational environment. An awareness of the necessity for identifying students who are at-risk or already failing heralds the implementation of progress monitoring systems that continuously survey acquisition of skills and development of subsequent academic competencies. Early literacy and mathematics skills are understood as essential prerequisites towards future academic achievement, emotional adjustment, and adult quality of livelihood.

Brief, reoccurring informal process assessment practices, such as DIBELS and Math Concepts and Applications, may yield a powerful mechanism to accomplish such progress monitoring and data based decision-making objectives. Previous quantitative approaches towards studying the outcomes of school-based data, however, were frequently plagued with methodological shortcomings and violations of statistical
assumptions. Advances in understanding nested or hierarchical organized data allows for analysis of data without many of these confounds.

The present study employed a longitudinal collection from 2002 to 2006 of informal DIBELS and *Math Concepts and Application* assessment results. Repeated measurement of a high-stakes measure, the Ohio Achievement Test subtests in reading and mathematics, were regressed onto informal math and reading assessments with various individual student-level predictor variables in a progressive sequence involving hierarchical linear models (HLM). The intent was to develop a cogent model of predicting high stakes achievement test performance as related to the above variables.

Results were significant for the usage of informal DIBELS measures to predict future high stakes achievement test performance but *MCA* partitioned only a minimal amount of variance in the regression equations. Despite *MCA*’s limited predictiveness, racial differences, special education participation, and overall school attendance were noted to affect mathematics high stakes test performance. In reading, significant performance differences were also noted in special education students. Such findings support previous literature on the utility of DIBELS but are discrepant for *MCA*. Still, these data remind educators of the saliency of early identification practices prior to the onset of reading or mathematics failure.
ACKNOWLEDGEMENTS

As an inveterate college student for more than twenty years, I have been influenced by legions of academic faculty, colleagues, family, and friends, across my scholastic, emotional, social, and professional development. Although there have been so many that this acknowledgement section should go on for chapters to pay adequate reverence, and it is probably lengthier than other’s offerings in a doctoral dissertation, the below offers a modest attempt to do justice given the longevity of my collegiate tenure.

First, I recall my undergraduate experience at Heidelberg College and the careful nurturance of Professors Wallace E. Dixon and Jeannine Studer, both of whom instilled a passion for understanding human behavior while facilitating my change from pre-medicine to psychology as a collegiate major. In my master’s training programs, my co-counseling program advisors, Professors David Brooks and Dale Pietrzak, both taught me how to look outside myself for the welfare of others while simultaneously never missing the opportunity to laugh at life’s permutations and myself, too. My second master’s thesis advisor, Professor Rhonda Richardson, permitted me the opportunity to complete a thesis and use it as a springboard towards future doctoral studies. Since taking doctoral coursework, I am deeply indebted to Professor Cathy F. Telzrow, who took a remarkable chance and admitted me to the doctoral program in school psychology. Although we frequently seemed at loggerheads with one another, I grew tremendously in my practitioner, scholastic, and maybe even social-emotional deportment through her influence. Sadly, because of many life events and arguably a profound lack of discipline...
on my behalf, I lost the opportunity to learn from her in the dissertation process when she retired (which hopefully my foibles did not hasten!). In addition, Professor Caven S. Mcloughlin has always been a resolute advocate for all his students, including myself. He reined me in when I was too askance, offered a solid voice of reason when I needed it, and of course espoused tremendous passion for the discipline of school psychology. Even today, I feel honored that he was willing to persevere with me throughout the storied conclusion of this dissertation when my absence of action and never-ending tales of woe should have caused him to abandon my plight long-before. Thank you, Dr. Mcloughlin; indeed, I know that I would not have my career or this manuscript without your professorial collegiality and mentorship. In my current dissertation committee, I am also appreciative of Dr. Frank Sansosti, with whom I came to celebrate his congenial yet practical council. I am much humbled by his support in this process even though I feel as if I have not had sufficient interaction to deserve his contribution. My friend, Dr. Jason Schenker, who has been not only a meaningful companion throughout much of my adult life (despite episodic nonsense I created for myself) but also truly taught me much more about my dissertation methodology and data analysis than I would have achieved otherwise. He altruistically gave robust support in the 11th hour when I needed to complete this dissertation project. Finally, many thanks to Dr. Manfred Van Dulmen, Assistant Professor, Kent State University Department of Psychology, for agreeing to serve as the graduate faculty representative and volunteering his time to support my dissertation defense process.
I am also privileged to share life with my partner, Holly J. Duff, who readily availed herself to this dissertation cause. She extracted interminable data points, coded data in boundless spreadsheets, cared for our home while I was preoccupied (or procrastinating on this dissertation), provided unwavering emotional support, and gave probably more than a few stern lectures about “finishing that dissertation, already.” Even when I started to believe I could not achieve it or otherwise should resign myself to failure, you were there. “Hols,” you were my pedestal, confidant, and sometimes the very necessary critical parent that helped me redirect my wayward efforts to this project. My other companion, Mr. Rolex, our Labrador retriever, was far more than the watchdog his namesake suggests. He was a constant source of comfort, laughter when he would dismantle all of my dissertation papers and supplies (even if I did not consider it as such at the time), and unwavering companionship during those long hours “daddy” was burning the midnight oil.

In the work dimension, perhaps no credit here is more significant than that due to Mr. Donald Saracusa, without whom, I would not have had access to this longitudinal dataset. His forethought in using informal measures to identify at-risk students was a decade sooner than most of our profession and his passion today in early identification for intervention remains an archetype for all school psychologists. In sharing a passion for distance running with Don, I wish you many more happy trails and cherish the road you allowed me to travel here! I wish to acknowledge my first doctoral supervisor, Dr. Richard Montagnese, for his years of clinical supervision and paternal guidance during
those crucial formative years of my pre-doctoral internship experience. As well, my former chief psychologist, Dr. Elizabeth Bard, who continuously supported me with friendship, kindness, and recurrent support as I progressed, in the slowest of trajectories. Finally, Dr. Perry J. Clark, for his pragmatic and trenchant feedback about work and life as well as his and none-too-subtle directive to finish my dissertation before another chief psychologist retires.

Separately, I owe a debt of gratitude to my current colleagues for allowing me access to these data and permission to conduct this study, including Dr. Patricia Agble, Mrs. Judy Gray, and the district administrative team. Dr. Robert J. Kubick, Jr. deserves an honorable mention for serving as my dissertation editor and sounding board throughout our shared doctoral journey together; even if you did finish your dissertation before me! Finally, no reflection on work would be complete without noting our fraternity of colleagues in the Child Study Department and the “Five Figure Income Boys Club” (FFIBC) for their friendship, congenial presence, and more than a few luncheon buffets to the detriment of my midsection.

This work is dedicated to the memory of my maternal uncle, Raymond, and maternal grandmother, Gertrude. Without both, I never would have been able to attend or continue higher education. My deepest regret is that I have lost both before I share this accomplishment with them – you are both much loved and deeply missed. With this conclusion, my formal college journey may be over, but I know learning about myself, life, the world, and becoming a more perspicacious psychologist has hardly begun…. 
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xi</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xii</td>
</tr>
<tr>
<td>CHAPTER 1: REVIEW OF THE LITERATURE</td>
<td>1</td>
</tr>
<tr>
<td>Twenty-First Century Educational Practice</td>
<td>1</td>
</tr>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Educational Accountability and Reform</td>
<td>3</td>
</tr>
<tr>
<td>Historical Context</td>
<td>3</td>
</tr>
<tr>
<td>Elementary and secondary education act of 1965</td>
<td>6</td>
</tr>
<tr>
<td>Individuals with disabilities education act of 2004</td>
<td>7</td>
</tr>
<tr>
<td>Reading &amp; Mathematics in Educational Accountability and Reform</td>
<td>8</td>
</tr>
<tr>
<td>Reading Achievement</td>
<td>8</td>
</tr>
<tr>
<td>Reading as educative process</td>
<td>9</td>
</tr>
<tr>
<td>Assessment of reading skills</td>
<td>10</td>
</tr>
<tr>
<td>Mathematics Achievement</td>
<td>12</td>
</tr>
<tr>
<td>Mathematics as educative process</td>
<td>13</td>
</tr>
<tr>
<td>Assessment of mathematical skills</td>
<td>14</td>
</tr>
<tr>
<td>Valuing Mathematics and Reading Achievement in NCLB</td>
<td>15</td>
</tr>
<tr>
<td>Measuring What Matters: Determination of Educational Success</td>
<td>16</td>
</tr>
<tr>
<td>Educational research involving VAM &amp; HLM: New paradigms for measuring success</td>
<td>16</td>
</tr>
<tr>
<td>Multilevel modeling</td>
<td>18</td>
</tr>
<tr>
<td>The Current Study</td>
<td>20</td>
</tr>
<tr>
<td>Research Questions</td>
<td>22</td>
</tr>
<tr>
<td>Level-one research questions &amp; hypotheses</td>
<td>23</td>
</tr>
<tr>
<td>Level-two research questions &amp; hypotheses</td>
<td>24</td>
</tr>
<tr>
<td>Slope related questions</td>
<td>24</td>
</tr>
<tr>
<td>Intercept related questions</td>
<td>26</td>
</tr>
<tr>
<td>CHAPTER 2: METHODOLOGY</td>
<td>29</td>
</tr>
<tr>
<td>Research Design</td>
<td>29</td>
</tr>
<tr>
<td>Participants</td>
<td>30</td>
</tr>
<tr>
<td>Instruments</td>
<td>36</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Level-1 OAT-M and Slopes of Grand Mean Centered MCA</td>
<td>78</td>
</tr>
<tr>
<td>2.</td>
<td>Level-1 OAT-R and Slopes of Grand Mean Centered DIBELS</td>
<td>78</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. School Building Demographic Characteristics</td>
<td>31</td>
</tr>
<tr>
<td>2. Index Values, DIBELS ORF and OAT-R Administration Sequence</td>
<td>34</td>
</tr>
<tr>
<td>3. Index Values, MCA and OAT-M Administration Sequence</td>
<td>35</td>
</tr>
<tr>
<td>4. Originally Proposed Longitudinal Sample Attrition by School Building</td>
<td>59</td>
</tr>
<tr>
<td>5. Descriptive Composition of Sample by Assessment Content Area</td>
<td>61</td>
</tr>
<tr>
<td>6. Descriptive Statistics of Sample’s Attendance by Academic Content Area</td>
<td>62</td>
</tr>
<tr>
<td>7. Descriptive Statistics for OAT Reading Subtest Standard Scores by Grade, n = 71</td>
<td>63</td>
</tr>
<tr>
<td>8. Descriptive Statistics for OAT Mathematics Subtest Standard Scores by Grade, n = 91</td>
<td>64</td>
</tr>
<tr>
<td>9. Descriptive Statistics for Spring DIBELS Oral Reading Fluency by Grade, n = 71</td>
<td>65</td>
</tr>
<tr>
<td>10. Descriptive Statistics for Spring Math Concepts and Applications by Grade, n = 91</td>
<td>65</td>
</tr>
<tr>
<td>11. HLM Results for Unconditional Model: OAT Math</td>
<td>84</td>
</tr>
<tr>
<td>12. HLM Results for Unconditional Model: OAT Reading</td>
<td>85</td>
</tr>
<tr>
<td>13. HLM Results for Conditional Model: OAT Math</td>
<td>88</td>
</tr>
<tr>
<td>14. HLM Results for Conditional Model: OAT Reading</td>
<td>89</td>
</tr>
</tbody>
</table>
15. Results of Full Conditional Hierarchical Linear Model: OAT Mathematics Subtest .......................................................... 95

16. Results of Conditional Full Hierarchical Linear Model: OAT Reading Subtest .............................................................. 99

B1. Tabular Presentation of Level-1 Variables in Analysis ............... 132

B2. Tabular Presentation of Level-2 Variables in Analysis ............... 133
CHAPTER 1: REVIEW OF THE LITERATURE

Twenty-First Century Educational Practice

Introduction

Contemporary educational practice during the twenty-first century has been saliently dominated by two polarizing reforms in the scholastic literature, legislative enactments, and popular media: (a) school choice as epitomized by the charter school movement and (b) testing and accountability as mandated by the eponymously entitled, *No Child Left Behind Act of 2001* (Good, Simmons, & Kameʻenui, 2001; Petrilli, 2008). These two reformation efforts have collectively helped to redefine the overall culture and politics of the American educational climate. Systematically speaking, each of the above topical points feed into one another recursively, pursuant to growing public concern about the success (or speculative failure) of American public schools, within themselves and in comparison to the larger global community. Historic efforts in education originally focused upon ensuring equal access and opportunity for all students regardless of gender, race, and disability status. Today’s emphasis has left these child-centered, and arguably to some, humanistic values, instead highlighting educational outcome products as markers of success (Braden & Tayrose, 2008). Establishment of statewide academic content standards, alignment of various assessments to such standards, implementation of accountability measures, and support of teacher quality through value-added assessment and legislative action are all seen as catalyzed by this educational improvement.
movement. Indeed, the apparent ultimate goal of all these actions is to improve student achievement, close achievement gaps, and of course, in the parlance of law, originally borrowed from the Children's Defense Fund, “leave no child behind” (U.S. Department of Education, 2004).

Central to this era of education reform and accountability is the importance of universal screening, understood as a “systematic assessment of all children within a given class, grade, school building, or school district, on academic and/or social-emotional indicators that the school personnel and community have agreed are important” (Ikeda, Neeseen, & Witt, 2008, p. 103). A linkage can be established between these large-scale testing efforts and the totality of educational reform to provide a macro-level picture of both the systemic and individual, or child-level, scholastic health in their quest to fulfill the legal dictates of the ongoing transformational process. Under both the mandates of No Child Left Behind [NCLB] and the Individuals with Disabilities Education Improvement Act [IDEIA] (Individuals with Disabilities Improvement Act of 2004, 2004), schools are now charged with demonstrating that students with and without disabilities are accessing and benefiting from the general educational curriculum. Born of these screening efforts is also a multi-tier approach towards student assistance whereby pupils are screened and subsequently moved through a multi-tiered model of intervention that serves both an immediate accountability purpose but also ensures connection to more targeted forms of intervention as appropriate (Tilly, 2008). Although many psychoeducational tools are available to screen students on various academic and behavioral dimensions, considerable
attention is given to the primacy of early reading and mathematics achievement due to the predictive power each holds in future academic and emotional success among school children (National Assessment of Educational Progress, 2005). Indeed, based on meta-analytic investigation, few predictors of future academic achievement are as strong as early reading and mathematics competency (e.g., Torgersen, 1998a; Duncan et al., 2007).

The remainder of this present chapter is dedicated to examining each of the above discussion points with the intent to describe how the present educational Zeitgeist has arrived within a concurrent obligation of demonstrable success, across specific mathematics and reading academic content areas, to the larger issue of value-added or hierarchical models. Because individual student success is also predicated on high-stakes testing performance through successive terms of recursion involving these educational reform practices, the utility of early reading and mathematics screening as a predictor on such larger-scale achievement outcome measures will be considered through a quantitative exploratory investigation using multilevel models.

Educational Accountability and Reform

Historical Context

American education is not articulated as a function of the federal government under the United States Constitution per se, leaving its overall regulation, funding, and daily activity reserved to individual state discretion (Cambron-McCabe, McCarthy, &
Thomas, 2004). Federal involvement in the individual state governance of education is chiefly through congressionally supported financial appropriations accepted by each state. These federal funds carry distinct stipulations that serve as a de facto assertive power over individual states’ ability to individually govern their respective educational systems.

Beginning with the publication of *A Nation at Risk: The Imperative for Education Reform* (National Commission on Excellence in Education, 1983), a movement was established for public accountability in schools that altered the educational landscape through the present day. Prominence was redirected from historic interests in equal educational opportunities with respect to gender, race, social-economic status, and disability condition, to an arrival of outcome products, such as large-scale achievement test performance by individual students. Language discussing standards-based education arrived during the latter 1980s, with the intent to a) institute a consistent model of what students should know and be able to do; b) align measurements consistent with those standards; and c) emphasize teacher improvement at the state-level (Braden & Tayrose, 2008). Interested readers may reference Falkenberg’s (1996) cogent chronology of the literature related to the emergence of standards-based education practices through examination of grade retention and promotion practices. Interestingly, despite the fact that such efforts were experienced as novel or revolutionary at their time of introduction, archival literature suggests that Starch (1915) already pre-dated this momentum nearly a century ago, championing that public schools require standards because it would allow a
“qualified person” to enter any classroom and determine whether pupils are “…up to the standard, whether they are deficient, and in what specific respect” (p. 15).

The entrance of educational standards (q.v., what do I teach?) was not considered sufficient in facilitating school improvement, however, even though many educators lauded these efforts because outcomes were not previously articulated. Instead, to assess the movement towards school improvement, various evaluative measures were developed that sought to clarify whether such standards were being achieved. In time, these measures were given the moniker “high-stakes” tests, because they brought serious ramifications to states, schools, and recently, even individual teachers and students.

Federal legislative efforts have come to mandate statewide high-stakes achievement tests, such as the Ohio Achievement Test (Ohio Department of Education, 2009), as well as prescribe various consequences based on those test results. Individual student consequences such as grade promotion or attainment of a high school diploma are not, however, established by any federal law or policy, but instead are added by state departments of education in hopes of stimulating student performance. Educational standards are generally classified into two groups: (a) content and proficiency standards that establish what students are supposed to know and exhibit; and (b) “opportunity to learn” standards that describe various supports needed by children to acquire high levels of academic proficiency (Braden & Tayrose, 2008). Federal law requires each state to have content and proficiency standards, yet does not enforce or otherwise monitor students’ opportunity to learn the same. In addition, marked criticism of this standards
movement has resulted, because high-stakes achievement measures disproportionately affect individual students over the larger educational system within which they attend (Doran & Fleischman, 2005). Clark (1997) bemoaned the arrival of these high-stakes measures, noting that they have “…created a new population of at-risk students” (p. 144). Despite this acrimonious chorus, standardized testing is not new to public education in the United States. Achievement tests have a long historic lineage of measuring students’ progress, sorting students by ability or disability, and tracking students towards various career paths (Gallagher, 2003).


**Elementary and secondary education act of 1965.** ESEA serves as the foremost legislative act affecting American education, resulting in excess of $130 billion per year spent towards supporting the educational system (Braden & Tayrose, 2008). A recent reauthorization of this legislation in 2001, termed “No Child Left Behind” [NCLB] added the additional expectation that individual states, districts, and local schools are now responsible for student achievement outcomes - a pointedly neglected aspect of the original 1965 ESEA legislation. NCLB further fundamentally changed the educational
landscape by requiring accountability for results, emphasizing scientifically based research, expanding parental options, and expanding flexibility and local control (U.S. Department of Education, 2004). Individual schools and school districts are expected to demonstrate adequate yearly progress [AYP] towards specific objectives and define what is meant by “proficient” in various academic measurements, especially math and reading. To ensure states are facilitating achievement, their results must be compared to a national benchmark - the National Assessment of Educational Progress [NAEP] - that compares fourth and eighth grade student achievement in math and reading domains. Despite the heightened emphasis on reading and mathematics achievement in NCLB, individual states appear to have set lower standards than NAEP and public schools are consistently noted to score below comparable private schools (Lubinski & Lubinski, 2006).

**Individuals with disabilities education act of 2004.** The Individuals with Disabilities Education Act of 2004 (IDEA), represents the reauthorization of the Education for All Handicapped Children Act (EHA; Education for All Handicapped Children, 1975) originally passed in 1975. IDEA serves as the primary federal legislation ensuring students with various disability conditions have equal access to a free and appropriate education [FAPE]. As related to educational reform, IDEA also attempts to align special education legislation with NCLB accountability requirements for the first time in history. Specifically, children with disabilities are now expected to achieve the same levels of academic proficiency as their non-disabled peers and their achievement data are required to be included in individual state accountability systems just as
non-disabled peers (Braden & Tayrose, 2008). In situations where children have
disabilities that preclude participation in traditional state assessment systems, modified
curricular and alternative assessment options are available. Interestingly and perhaps
unintended with IDEA legislation, Kubick (2007) noted that aspects of IDEA involving
behavioral supports for students with disabilities may also yield increased outcomes on
high-stakes math testing, in certain circumstances, compared with historic models of
behavioral discipline efforts that may have attenuated performance in that content area.

**Reading & Mathematics in Educational Accountability and Reform**

As previously reviewed above, the passage of NCLB shepherded a new era where
reading and mathematics proficiency was not only expected across all students enrolled
in public school but considered a crucial aspect of emergent educational reform practices
to ensure individual states are meeting NAEP comparison data. In this next section, the
importance of these academic content areas will be explored alongside commonplace
universal screening methods to document student progress in such domains.

**Reading Achievement**

Across the nation, there is growing awareness of the importance of, and dividends
paid through, early reading success, as well as the adverse consequences of early reading
impairment (Torgensen, 1998a; Good, Simmons, & Kame'enui, 2001). Recent data
suggest that more than 1.5 million students are at-risk for not reading on grade level and
require some degree of intervention to adhere to legislative requirements of NCLB’s
reading dictates (Roehrig, Petscher, Nettles, Hudson, & Torgesen, 2008). In one study, more than 36% of American fourth grade students had not achieved a basic level of reading proficiency, leading to significant social and economic disadvantage for these pupils in subsequent years (Hosp & MacConnell, 2008; National Assessment of Educational Progress, 2005). Torgesen (1998b) notes that children who start their educative process as poor readers, rarely, if ever, catch up to their peers. Based on this sentiment, Torgensen and colleagues (e.g. Torgensen, Wagner & Rashotte, 1994; Torgensen, Wagner & Rashotte, 1997), pointedly admonish that the first grade student who struggles in reading is likely to exponentially accumulate monumental and pervasive academic deficits throughout the remaining years of schooling. These cumulative effects yield students who harbor negative attitudes towards reading as a whole, missed opportunity for developing reading comprehension strategies throughout formal schooling, and ultimately, become individuals who practice the skill of reading substantively less than their peers (Torgensen, 1998b).

**Reading as educative process.** Efforts to define early identification procedures for reading difficulties have a long and rich history in educational and psychological measurement that far exceed the capacity to synthesize in a single literature review (Speece, 2005). In general, the broad process of reading may take several forms, inclusive of decoding, analogizing, or prediction to access and comprehend unfamiliar content. Readers can determine novel words through a process of forming connections between graphemes and phonemes to bond spellings to their pronunciations and
meanings in memory. Phonemic awareness results from knowledge of the alphabetic system, which serves as a powerful mnemonic in future word decoding activities (Ehri, 2005). Reading fluency, operationalized as the speed and accuracy with which text is read orally, is considered critical to skilled reading, given its correlational connection to comprehension (Speece & Ritchey, 2005). These developmental sequences of reading skill acquisition are noted to be relatively universal in Western societies with formalized educational structures such as noted in Kudo and Bazan’s (2009) investigation of beginning reading skills among Peruvian schoolchildren.

**Assessment of reading skills.** For individual states and local schools to receive Title I funding from NCLB, applicants are expected to incorporate assessment and scientifically based research into improving reading outcomes. Underlying this law is the assumption that such activities, when linked to instructional planning, will facilitate improved teaching and learning of reading (Collins, 2007). The current educational climate challenges teachers with finding ways to effectively document student responsiveness to interventions and track progress towards important achievement outcomes. In reading, commonly used formative assessment approaches involve evaluating how a student has mastered or attained a given instructional objective as represented in the overall curriculum.

Stepping away from such curriculum-based measures, Dynamic Indicators of Basic Early Literacy Skills [DIBELS] has enjoyed widespread support in identification of
children having trouble in the early acquisition of basic early literacy skills from kindergarten to sixth grade (Kaminski, Cummings, Powell-Smith, & Good, 2008). DIBELS was based on measurement procedures established for curriculum-based measurement [CBM] by Deno and colleagues in the 1970s and 1980s, but subsequently modified to be generic and draw content from multiple sources rather than any specific curriculum. Elements of DIBELS include initial sound fluency, phoneme segmentation fluency, nonsense word fluency, oral reading fluency, word use fluency, and retell fluency, with each basic early literacy skill administered at specific grade levels. Composite results of DIBELS literature suggest that these measures can discriminate those students who will pass high-stakes large-scale reading achievement tests during the third grade (e.g., Good, Simmons, & Kame'enui, 2001).

Torgensen (1989, 2001) supports these “intrinsic process approaches” to reading skill assessment as a more potent methodology towards the practical differential diagnosis of reading learning disabilities in comparison to the traditional discrepancy models between intellectual ability and predicted achievement levels. While the utility of intellectual assessment in reading and overall learning disability diagnosis remains controversial in the literature (e.g. Siegel, 1988), the onset of the response to intervention movement (e.g. Fuchs & Fuchs, 2008) and eschewal of individualized norm-reference standardized evaluation has gained momentum given the power of many informal measures to predict academic performance. Through examination of the various components of reading skills, such as employed with DIBELS and oral reading fluency,
for example, the saliency of the IQ score within the special education identification process is reduced, even though the utility of intellectual ability remains relevant for the purposes of scientific research on learning disabilities. Direct assessment, thusly, represents a vehicle towards differentiating children with true specific learning disabilities over those with environmental and constitutional needs such as a lack of educational opportunity or limited motivation to learn (Torgenson, 2001).

Mathematics Achievement

Similar to the saliency of reading as a predictor of overall school success, mathematics is coequally entrenched and described as a requirement of most jobs and successful independent living within the community (Patton, Cronin, Bassett, & Koppel, 1997). Mathematical competency as a national educational goal was emphasized both in NCLB and through the formation the National Mathematics Advisory Panel (Kelley, 2008). Despite these efforts, conceptual success in mathematics is defined in varying ways by different professional sources. Kilpatrick, Swafford, and Findell (2001) argued that math proficiency includes five interconnected strands: (a) conceptual understanding, (b) procedural fluency, (c) strategic competence, (d) adaptive reasoning, and (e) productive disposition. In contrast, the National Council of Teachers of Mathematics [NCTM] described the two basic categories of math as mathematical reasoning and specific math content (National Council of Teachers of Mathematics, 2000). Regardless of definitional resource, mathematical skill deficits appear widespread in the preschool
through twelfth grade American educational population. Recent data indicated that 21% of fourth graders and 32% of eighth graders continue to perform below basic levels of grade expectancy proficiency (National Assessment of Educational Progress, 2005). Additionally, the prevalence of math learning disabilities have been reported as high as 6.5% (Gross-Tsur, Manor, & Shalev, 1996) and extending beyond the roughly 5% range expected by epidemiological data (Fuchs & Fuchs, 2005; Kelley, 2008).

Mathematics as educative process. The curricular and instructional debate concerning mathematics notwithstanding, the educative process of mathematical problem solving is commonly organized around conceptual, factual, procedural, and application/problem solving elements (Kelley, 2008). Conceptual knowledge is stated as the primary goal of math instruction by NCTM because progression through mathematical curricula presupposes an architectonic acquisition of math skill development. Strategic or procedural knowledge of mathematics refers to the general procedures followed to combine subtasks into larger tasks, with math exhibiting a sequential series of steps towards solving computational, application or word problem examples. A potent challenge facing math teachers is the frequently observed phenomenon of students demonstrating procedural knowledge in the absence of conceptual knowledge. For example, a student may arrive at the correct answer in solving a division problem involving fractions through rote memorization of the sequential steps, but otherwise fail to understand the bigger conceptual matter of partitioning out fractional components of an integer. Factual knowledge in mathematics explains exclusive elements
of information that could include math-specific vocabulary or specific details necessary to understand the Gestalt of a problem. Finally, application knowledge alternates depending on referential source and adoption of local math curriculum as either knowledge domains or knowledge display. The essential feature of this last element refers to the overall totality in which students learn math and apply those specific skills to the bigger picture of their situation.

**Assessment of mathematical skills.** Evaluation of student mathematical skills involves a problem-solving process to determine qualitatively what skills are deficit and how best to intercede (Kelley, 2008). Mathematical skills are not present in isolation and may depend, in part, on reading oral fluency achievement. Indeed, a deficit of reading also negatively influences mathematical performance as indicated by data collected in Washington state (Hirsch, 2006). Ambitious NCTM standards advocate for assessing the “full mathematical power of students” beyond their simple capacity to perform routine computational procedures and isolated skills as established in individual state academic content standards (Boyd, 2008; National Council of Teachers of Mathematics, 1995).

Best practices in the school psychological literature affirm the process approach towards assessment, suggesting that curriculum based evaluation [CBE] involve components of survey level math assessments [SLAs], general outcome measures [GOMs], and skills-based measures can answer the foundational questions of what to teach and how to teach it. The general practice, similar to other CBE approaches in
reading or writing, involves problem identification, hypothesis generation, problem analysis, plan development, and plan evaluation (Kelley, 2008). Separately, progress monitoring of reading and mathematical skills move existing educational practice and assessment methodologies towards a multi-tiered prevention system for all students. In such an approach, a generalized response to universal interventions is established. Should students be unsuccessful with these lower-tier interventions, targeted interventions are hypothesized and intensive individualized instruction is identified for those progressing the least. Upon reaching the final tier of intervention, consideration for individual psychoeducational evaluation or return to previous tiers of intervention can be assessed (Fuchs & Fuchs, 2008).

Valuing Mathematics and Reading Achievement in NCLB

Accountability and testing has taken primacy in the NCLB plan to improve public schools (Boehner, 2002). With the reauthorization of ESEA, states that elect to receive those categorical funds are empowered to create and implement “challenging standards,” specifically in the areas of reading and mathematics, administer annual testing to all students in grades 3-8, and ultimately have all students proficient by 2014 regardless of group affiliation (Braden & Tayrose, 2008; U.S. Department of Education, 2004). Consequently, the urgency in identifying those students who are at-risk for reading or mathematical difficulty has taken on additional necessity to ensure they are achieving and facilitating the generalized success of each state and respective school district.
Determination of a successful educational program, however, requires new methods for evaluating the effectiveness of teachers and schools, and such methods appear beyond simple administration of large-scale achievement measures as codified in NCLB (Doran & Fleischman, 2005; Peterson & Hess, 2008).

**Measuring What Matters: Determination of Educational Success**

**Educational research involving VAM and HLM: New paradigms for measuring success.** The current educational climate features increased scrutiny of the accountability measures of teachers and schools, largely as a function of NCLB legislation that requires state-level departments of education to establish a system of rewards and punishments for districts and schools accounting for student achievement (Rubin, Stuart, & Zanutto, 2004). The consequence of this culture has lead to an increase in large-group, high-stakes testing with explicit publication of school rankings and test results. Estimation of school and teacher effects have been accomplished through a variety of statistical models, typically described as “value added measures” [VAM] in the educational literature (Sanders, 2000).

VAM measures gained popularity after Sanders (2000) developed a statistical methodology he adapted from agricultural research. Because of arguments in the disparities of school funding, Sanders implemented the Tennessee Value Added Assessment System [TVAAS], which employed a longitudinal database that linked student test scores to their teachers and to their specific schools (Doran & Fleischman,
2005; Sanders, 2000). VAMs vary from uncomplicated fixed effects models (e.g., Tekwe, et al., 2004) to complex, multivariate, longitudinal mixed-models with either test scores or increases in test scores as outcome products (e.g., McCaffrey, Lockwood, Koretz, Louis, & Hamilton, 2004). Generally, these latter models incorporate various statistical parameters for school/teacher, student-level, classroom-level, and school-level covariate effects, with the opportunity to complete intra-class correlation among outcomes for students in the same class. The goal of VAM literature is to estimate causal effects or how a specific teacher or school has “added value” to their students’ test scores (Rubin, Stuart, & Zanutto, 2004).

Much discussion abounds on VAMs and their impact on education. Proponents argue that such measures give a powerful diagnostic tool for measuring the effect of pedagogy, curricular, and professional development on academic achievement, giving American education a new unbiased model for determining accountability as mandated under NCLB (Hershberg, Simon Adams, & Lea-Kruger, 2004). Detractors of the movement cite that the VAMs make unequal comparisons across schools and teachers who may not have demographically similar populations of students, thereby causing unnecessary scrutiny towards schools and individual teachers. Because of the wide variation in computing VAMs, the issue of missing data has also brought criticism as some models treat the those data points as random or assuming growth would be the same as other peers, causing skew in the results (Olson, 2004). The computations of such models also are seen as particularly complex, requiring educational units to collaborate
with various academic or consultative firms to handle the necessary machinations of the statistical calculations (Doran & Fleischman, 2005). Regardless of the perspective on the tool, it also highlights the necessity of having multilevel models in educational research that attend to inherent violations of past methods, particularly within nested designs, repeated measures of such nested designs, and intra-class correlational effects.

**Multilevel modeling.** A recurrent methodological criticism in educational research over the past decade has been the failure of many quantitative studies to attend to the hierarchical, multilevel nature of general educational field data (Raudenbush, 1988). Traditionally, research involving multilevel data was accomplished through ordinary least squares [OLS] and analysis of covariance [ANCOVA], neither of which are appropriate for datasets that are nested in nature due to the presence of within and between group variance (Pedhazur, 1997). For example, OLS fails to account for contextual effects, which eliminates the group characteristics, potentially causing information from one level to be missed in the broad statistical calculations. ANCOVA suffers a similar limitation, in that it does not account for intra-class correlative effects (Kreft & de Leeuw, 1998). When the nested structure of school-based data, in particular, are not accounted for, two serious methodological flaws almost inevitably result: (a) aggregation bias (Cronbach & Webb, 1975); and (b) misestimated precision (Raudenbush, 1988). Aggregation bias creates an increase in Type I error by artifically inflating the likelihood that group differences exist where none are present (Snijders & Bosker, 1999). Unlike hierarchical linear modeling [HLM], which calculates error at each
level, misestimated precision occurs when variance at each level is not estimated correctly within a unidimension or un-leveled statistical analysis (Boyle & Willms, 2001).

In response to these criticisms, several methodologists have created statistical multilevel (q.v., contextual, multilinear, mixed linear, random coefficient) models that seek to: (a) enable researchers to formulate and test for processes occurring within and between educational units (Raudenbush, 1988); and (b) provide methods to specification of appropriate error structures, including random intercepts and random coefficients. Furthermore, unbalanced designs are now permitted with the introduction of asymptotically efficient estimates of the variances and covariances of random effects. Raudenbush (1988) generally used the terminology of “hierarchical linear models [HLM]” because such statistical approaches use hierarchically structured or nested data that generally operationalize the statistic as parameters at the lower level of aggregation. From there, these parameters are presumed to vary as a function of the computed coefficients at the next higher level of the hierarchy.

Hierarchical linear modeling techniques have gain considerable popularity for educational research that seeks to partition out the contribution of various factors or components in student educational performance. For example, Taylor et al. (2002) employed HLM to examine how various programmatic and classroom instructional factors relate to reading achievement. Their results complimented other studies noting
that enhanced reading achievement is both related to what and how teachers teach, specifically as related to various aspects of reading skills such as letter naming, rhyming, phonemic awareness, and fluency, consistent with DIBELS. Equally so, their methodology reveals how HLM can be effectively used to examine the contextual effects of specific classroom and teacher characteristics upon the larger outcome of mean academic reading performance as a research methodology.

The Current Study

The purpose of this study was to examine elementary student performance on the Ohio Achievement Test [OAT] reading and mathematics subtests, reflecting a high-stakes achievement measure, and to establish a predictive and discriminating model of individual student performance using informal DIBELS oral reading fluency [ORF] and informal math concepts and applications [MCA] measures. Data were obtained through a longitudinal dataset, composed of all students attending two elementary buildings, grades one through five, between 2001-2005 academic years for the DIBELS/MCA informal measures, and 2003-2006 for the OAT reading and OAT mathematics subtest results. The study also investigated existing archival records of these students to gather yearly attendance rates, designation as a student receiving special education services as well as demographic information including race and gender. Using multilevel modeling techniques, specifically hierarchical linear modeling [HLM], the development of a two-level model was proposed. On the first level, Ohio’s high-stakes
large-scale achievement measure, the Ohio Achievement Test [OAT], was regressed onto the independent variables of DIBELS ORF and Math Concepts and Applications [MCA] through multiple multivariate regression techniques to determine how these informal measures partitioned the variance related to performance on the OAT. On the second level, coefficients from the first-level analysis became the dependent or response variables that were regressed onto various demographic characteristics of the students, including race, gender, attendance, and special education services. These second-level equations both predicted OAT subtest standard scores through their regression equation intercepts and the relationships between each of these latter individual student variables as a function of their contribution to the variance attached to DIBELS ORF and MCA as it predicted OAT reading/math subtest performance. Consistent with the recommendation of Singer and Willet (2003), the multilevel models were estimated through a progressive building process so that the final model, with adequacy of fit to the data established, could address the research questions of interest, theoretical orientation or dictates of the study, and meet the demands of the available data. The intent of this study was multi-fold: (a) to demonstrate the efficacy of early informal measurement tools in predicting future high-stakes test performance on the OAT; (b) to predict whether any individual demographic student variable correspondingly predicted OAT performance when DIBELS ORF and MCA were administered in previous years; (c) to distinguish if DIBELS ORF and MCA were more appropriate screening measures for predicting OAT achievement across one or more micro units of students after controlling for race, gender,
and special education services; and (d) to predict strength of relationships from DIBELS and MCA as related to OAT performance. Such an investigation fits the contemporary educational culture of NCLB reforms (Good, Simmons, & Kame'enui, 2001; Petrilli, 2008) with emphasis on improving all student achievement, particularly within reading and mathematics (U.S. Department of Education, 2004). Moreover, this study supported an approach for universal screening consistent with best practice recommendations to support student learning in the zeitgeist of accountability and educational reform (Braden & Tayrose, 2008; Fuchs & Fuchs, 2008).

**Research Questions**

The research questions of this study were organized by the various levels of hierarchical models purported to be developed using the HLM methodology in the subsequent discussion. Level-one research questions simply involved OAT reading and math subtest standard scores as response variables regressed onto repeated measures of DIBELS ORF and MCA paired at the individual student unit according to the informal measure raw score and the subsequent year’s OAT subtest standard score. Level-two research questions contributed towards building the overall hierarchical linear model and answered specific predictive and relational questions based on the coefficients determined in the first level research questions.
Level-one research questions and hypotheses.

Question L1-1: Is there a significant predictive relationship between DIBELS oral reading fluency [ORF] and Ohio Achievement Test [OAT] reading subtest performance for elementary students?

Research hypothesis L1-1: There is a significant positive predictive relationship between a previous year’s DIBELS ORF measurement and a subsequent year’s OAT reading performance, \( (p<.05) \).

Null hypothesis L1-1: There is no significant predictive relationship between previous DIBELS ORF measurement and future OAT reading performance, at least at a statistically significant level, \( (p \geq .05) \).

Question L1-2: Is there a significant predictive relationship between informal math concepts and applications [MCA] and OAT math subtest performance for elementary students?

Research hypothesis L1-2: There is a significant positive predictive relationship between a previous year’s MCA measurement and a subsequent year’s OAT math performance, \( (p<.05) \).

Null hypothesis L1-2: There is no significant predictive relationship between previous MCA measurement and future OAT mathematical performance, at least a
Level-two research questions and hypotheses.

Slope related questions.

Question L2s-1: Does the moderator variable of attendance rates for individual students influence the strength of DIBELS/MCA relationship with OAT reading/math subtest performance after controlling for gender, race, and special education participation?

Research hypothesis L2s-1: There is a strong moderator effect between the relationship of DIBELS/MCA and those students who have increased attendance, after controlling for race, gender, and special education participation, \( (p < .05) \).

Null hypothesis L2s-1: There is no significant moderating relationship between DIBELS/MCA and those students who have increased attendance, after controlling for race, gender, and special education participation \( (p \geq .05) \).

Question L2s-2: Does the moderator variable of student race influence the strength of DIBELS/MCA relationship to OAT reading/math subtest performance after controlling for attendance, gender, and special education participation?

Research hypothesis L2s-2: There is a strong moderator effect between the relationship of DIBELS/MCA and race, after controlling for attendance, gender, and special education participation, \( (p < .05) \).
Null hypothesis L2s-2: There is no significant moderating relationship between DIBELS/MCA and race, after controlling for attendance, gender, and special education participation, \((p \geq .05)\).

Question L2s-3: Does the moderator variable of gender influence the strength of relationship between DIBELS/MCA to OAT reading/math subtest performance *after controlling* for attendance, race, and special education participation?

Research hypothesis L2s-3: There is a strong moderator effect between the relationship of DIBELS/MCA and gender, after controlling for race, attendance, and special education participation, \((p < .05)\).

Null hypothesis L2s-3: There is no significant moderating effect between DIBELS/MCA and gender, after controlling for race, attendance, and special education participation, \((p \geq .05)\).

Question L2s-4: Does the moderator variable of participation in special education influence the strength of relationship between DIBELS/MCA with OAT reading/math subtest performance *after controlling* for attendance, race, and gender?

Research hypothesis L2s-4: There is a strong moderator effect between the relationship of DIBELS/MCA and special education, after controlling for race, gender, and attendance, \((p < .05)\).
Null hypothesis L2i-4: There is no strong moderating relationship between DIBELS/MCA and special education, after controlling for race, gender, and attendance, ($p \geq .05$).

**Intercept related questions.**

Question L2i-1: Does higher school attendance predict OAT reading/math subtest scores for students performing at the mean on DIBELS/MCA, *after controlling* for race, gender, and special education participants?

Research Hypothesis L2i-1: There is a significant positive predictive relationship between higher attending students and OAT reading/math subtest scores, after controlling for gender, race, and special education participation, ($p < .05$).

Null hypothesis L2i-1: There is no significant predictive relationship between high attending students and OAT reading/math subtests, after controlling for gender, race, and special education participation, ($p \geq .05$).

Question L2i-2: Is there a significant predictive relationship of student gender to OAT reading/math subtest performance, for students performing at the mean on DIBELS/MCA, *after controlling* for race, attendance, and special education participation?
Research hypothesis L2i-2: There is a significant positive predictive relationship of gender on OAT reading/math subtest performance after controlling for race, attendance, and special education participation, \((p<.05)\).

Null hypothesis L2i-2: There is no significant predictive relationship of gender on OAT reading/math subtest performance, after controlling for race, attendance, and special education participation, \((p \geq .05)\).

Question L2i-3: Is there a significant predictive relationship between race and OAT reading/math subtest performance, for students performing at the mean on DIBELS/MCA, under control for gender, attendance, and special education participation?

Research hypothesis L2i-3: There is a significant inverse predictive relationship between race and OAT reading/math subtest performance, after controlling for gender, attendance, and special education participation, \((p < .05)\).

Null hypothesis L2i-3: There is no significant predictive relationship between race and OAT reading/math subtest performance, after controlling for gender, attendance, and special education, \((p \geq .05)\).

Question L2i-4: Is there a predictive relationship between participation in special education and OAT reading/math subtest performance, for students performing at the mean on DIBELS/MCA, under control for gender, race, and attendance?
Research hypothesis L2i-4: There is a significant inverse predictive relationship amongst special education participation and OAT reading/math subtest performance, after controlling for race, gender and attendance, \( (p < .05) \).

Null hypothesis L2i-4: There is no significant predictive relationship amongst special education participation and OAT reading/math subtest performance, after controlling for race, gender, and attendance, \( (p \geq .05) \).
CHAPTER 2: METHODOLOGY

Research Design

Individual student data were obtained from a longitudinal sample of informal measurement and cumulative file data, available within the school psychological department of a large, urban school district in Northern Ohio. Because of the multilevel nature of data, hierarchical modeling level-1 equation variables included the standardized (converted Rasch ability) scores from the Ohio Achievement Test [OAT] subtests of reading and mathematics with DIBELS Oral Reading Fluency [ORF] and informal Math Concepts and Applications [MCA] raw scores. Specifically, dependent variables on the level-1 equations included the OAT standardized scores, obtained across between grades 3 to 6 for reading and grades 5-6 for math, paired with the independent variable of the previous year’s DIBELS/MCA spring scores by individual student. Because of this pairing method, four sets of reading DIBELS and OAT-R scores were available while only two sets of MCA and OAT-M scores were. Level-2 explanatory variables towards building the larger model included total attendance between years 2002-2006, race, gender, and participation in special education and related services, again, linked by the identification variable of individual student number.
Participants

Participants in this study attended a large, urban school district in Northern Ohio composed of 7 high schools, 10 middle schools, and 34 elementary schools. In the longitudinal period of this investigation, three buildings were closed separately due to district efforts to “right size” their building infrastructure presence. The average daily enrollment of this district during academic year 2007-2008 was approximately 25,000 students.

This average daily membership [ADM] figure exhibited moderate decline from a high of approximately 29,000 students in the latter 1990s to the overall 2002-2006 longitudinal period in which study data were collected, consistent with trends of other urbanized environs experiencing “urban flight” (Clotfelter, 2001). The racial/ethnic distribution of the composite district student body can be disaggregated as follows: 48.0% Black, 42.4% White, 6.0% multiracial, 1.4% Hispanic, 2.0% Asian or Pacific Islander, and 0.1% Native American. 79.7% of students were identified as economically disadvantaged, 2.0% exhibited Limited English Proficiency [LEP], and 17.5% of the students were identified with one or more handicapping conditions under IDEA (Ohio Department of Education, 2008).

The total universe from which the current study’s sample was drawn included all students in grades 1-5 within two of the district’s 34 elementary school buildings. The two buildings specifically selected were staffed by the same veteran school psychologist,
with yearly trimester data collection efforts from a local university graduate program in school psychology, all of whom gathered informal reading and math data consistently across the years of the longitudinal investigation. Each building had unique demographic characteristics and is summarized in Table 1:

Table 1

*School Building Demographic Characteristics*

<table>
<thead>
<tr>
<th>Building Name</th>
<th>Average Daily Membership</th>
<th>Percentage of Student Population</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Black</td>
</tr>
<tr>
<td>“A”</td>
<td>251</td>
<td>70.5</td>
</tr>
<tr>
<td>“B”</td>
<td>338</td>
<td>45.7</td>
</tr>
</tbody>
</table>

*Note.* NC = Not calculated by Ohio Department of Education; when <10 students in a particular group

Although both samples were uniformly matched in their economic disadvantaged status, and absence of Limited English Proficiency [LEP] involvement, building "A" presented
itself as with a predominately Black American population while building "B" was roughly balanced in racial composition of White and Black American students. Some variability in these demographic trends was possible because parents or guardians individually self-selected their child’s racial membership during school enrollment, which does not yield a consistent approach towards racial membership identification.

Participants for the current study included all elementary students in grades 1-5, heterogeneously grouped as both general and categorically identified higher incidence special education recipients, nested in individual classrooms and in one of two school buildings as described above. Each student was administered the DIBELS ORF (Good & Kaminski, 2002) and Pro-Ed Monitoring Basic Skills Progress Math Concepts and Applications [MCA] (Fuchs, Hamlett, & Fuchs, 1998) as informal measurements of basic reading fluency and basic math computational and reasoning ability, respectfully. Standardized administration instructions to ensure consistency of measurement across the years of the investigation for these measures are presented in Appendix A for reference. Each measurement of the DIBELS ORF and MCA was repeated three times per academic year - fall, winter, and spring, with only the spring informal raw scores used in this present study. The rationale for restricting the study’s independent variables to only the spring informal administration is due to the fact that those scores should represent the highest level of reading and mathematical skill exhibition for a given year due to the aggregate linear nature of skill development in these academic content domains (Fuchs &
Fuchs, 2008). For the reader’s clarity, a tabular representation of the informal measure administration in comparison to the OAT administration sequence can be illustrated as follows; the index score refers to the specific data point identification for a given student with corresponding DIBELSxx and OATRyy values:
Table 2

*Index Values, DIBELS ORF and OAT-R Administration Sequence*

<table>
<thead>
<tr>
<th>Index Number</th>
<th>DIBELS &amp; OAT-R Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>DIBELS02 and OAT-R03</td>
</tr>
<tr>
<td>2</td>
<td>DIBELS03 and OAT-R04</td>
</tr>
<tr>
<td>3</td>
<td>DIBELS04 and OAT-R05</td>
</tr>
<tr>
<td>4</td>
<td>DIBELS05 and OAT-R06</td>
</tr>
</tbody>
</table>

*Note.* DIBELS<sub>xx</sub> refers to the individual participants’ grade of spring *DIBELS Oral Reading Fluency* administration while OAT-R<sub>yy</sub> represents the grade in which the *Ohio Achievement Test – Reading subtest* was administered to the same participant.
Table 3

*Index Values, MCA and OAT-M Administration Sequence*

<table>
<thead>
<tr>
<th>Index Number</th>
<th>MCA &amp; OAT-M Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MCA04 and OAT-M05</td>
</tr>
<tr>
<td>2</td>
<td>MCA05 and OAT-M06</td>
</tr>
</tbody>
</table>

*Note.* MCAxx refers to the individual participants’ grade of spring *Math Concepts and Applications* administration while OAT-Myy represents the grade in which the *Ohio Achievement Test – Math subtest* was administered to the same participant.

Students were also administered the Ohio Achievement Test [OAT] subtests in reading and mathematics each year, starting in 2003 as a state mandated requirement from the Ohio Department of Education [ODE]. Every participant in this study had equal opportunity to participate in the DIBELS ORF, MCA, and OAT, once this latter measure was adopted by ODE, throughout their entire elementary school career at these two school buildings. Even if a student were absence on the day of administration, ODE requires that subsequent re-administration of those measures are given, which helped to
ensure all students had those level-1 dependent variable points. The total universe of students, from which the study’s sample was drawn, \( N = 1455 \), was established by consolidating all students who would have had the possibility of participating in one or more administration of the DIBELS/MCA and OAT reading/math between the two school buildings and the period of the sample, 2002-2006. Given the repeated measures nature of the DIBELS and MCA, matching issues surrounding incomplete data, (e.g., one student missing a DIBELS ORF or an OAT subtest), do not represent the same potent degree of limitation as in other methodological approaches because HLM allows for missing data at the first level of analysis in a satisfactorily manner (Maas & Snijders, 2003). The interested reader is directed to Appendix C for additional discussion of data management issues with attention to understanding, addressing missing or incomplete individual student data, as well as data imputation using ML in the HLM software for the level-1 data. For students missing level-2 data, including race, gender, SPED status, or total attendance variables, those participants were removed from the present study as HLM cannot handle missing level-2 data, also described in Appendix C.

**Instruments**

**Dynamic Indicators of Basic Early Reading Fluency (DIBELS)**

DIBELS is a comprehensive set of various measurements intended to differentiate students who are struggling with various early reading concepts. Student data can be classified into three categories: at-risk, emerging, and established (Good & Kaminski,
The specific subtest of Oral Reading Fluency [ORF] was chosen for administration at the onset of this longitudinal data gathering experience by the building-level school psychologist and school district given its considerable empirical support as the most salient characteristic of reading competence (e.g., Adams, 1990) and as described in Chapter 1. Operationalized, oral reading fluency refers to the oral translation of text with speed and accuracy. ORF appears strongest amongst the various DIBELS subtests in discriminating basic reading skill processes with greatest growth occurring during the primary grades and a negative accelerating trajectory in latter elementary and middle school grades (Fuchs, Fuchs, & Hosp, 2001). In addition, Torgensen (1998a; 1998b) recommends process measures, such as oral reading fluency, as a quick and robust measure of general reading ability that can portend future reading and general composite academic achievement.

The measurement of DIBELS ORF requires the administration of three grade specific reading passages, each for a 1-minute interval, then an average score is taken in the number of words correctly read aloud. DIBELS ORF is administered three times per year - fall, winter, and spring - with established criterion cut-off points to qualitatively classify students in one of three areas of performance. Psychometric properties for this measure have been well established. Test-retest reliabilities for elementary students ranged from .92 to .97; alternate forms reliability of different reading passages drawn from the same level .89 to .94 (Tindal, Marston, & Deno, 1983). Criterion-related validity
studied in eight separate studies reported coefficients ranging from .52 to .91 (Good & Jefferson, 1998). Several existing studies have suggested the utility of employing oral reading fluency to predict later high-stakes statewide assessment performance in Michigan, Pennsylvania, Minnesota, Colorado, North Carolina, Arizona, and Ohio (e.g., Vander Meer & Stollar, 2005). In this current study, only DIBELS ORF repeated measurement scores were used out of the composite DIBELS subtest battery.

**Monitoring Basic Skills Progress: Math Concepts and Applications**

The *Monitoring Basic Skills Process* [MBSP] *Math Concepts and Applications* [MCA] are a research-based standardized set of measurement and evaluation procedures published commercially (Fuchs, Hamlett, & Fuchs, 1998). The measure is intended to focus intensively on the math progress of individual students who have identified learning problems and to evaluate formatively in order to improve student programmatic success. Such an evaluative approach is considered within the scope of best practices as formative evaluation through curriculum-based measurement in a process model (e.g., Boyd, 2008). In addition, these math curriculum-based measures appear to have predictive utility in performance of high-stakes statewide achievement tests (Fuchs & Fuchs, 2005). VanDerHeyden (2010) notes that process measures of mathematical skills, including combining computation, fluency, and application, are the best approach for identifying those students at future risk of mathematical failure and are recommended to be administered during primary years of schooling for identification and targeting.
students for intervention. On the measure, reproducible test page content varies by grade, with difficulty increasing over each grade level. Specific content items include counting, naming numbers, money, charts/graphs, fractions, decimals, word problems, applied computation, and measurement. Unlike DIBELS ORF for reading measurement, which displays a raw score increase over each year due to increasing reading fluency skills over time, the MCA is a separate and unique instrument by grade level with a specific raw score criterion range for a given grade.

Administration is standardized and requires a 6-minute timed interval where students are prompted to complete as many concept and applications as they can. The psychometric properties are acceptable, revealing adequate sensitivity and predictive screening utility (Cressey & Ezbeki, 2008; Fuchs, Hamlett, & Fuchs, 1998).

Ohio Achievement Test

Ohio began high-stakes testing of students in November of 1990. Students were then required to pass tests in reading, writing, mathematics, and citizenship to graduate and subsequently receive a standard Ohio public high school diploma (Boyd, 2008). Successive drafting of Ohio-specific content standards led to the development of the current Ohio Achievement Test [OAT]. The OAT represents a system of standards-based achievement tests administered in grades 3 through 8 in five subject areas: reading, math, science, social studies, and writing. The reading and mathematics subtests are administered annually in grades 3 to 8, science and social studies tests are administered in
grades 5 and 8, while the writing achievement test is administered in grades 4 and 7.

Unique to the reading OAT subtest is its administration twice a year during grade 3, with “proficient” level students excused from the latter spring re-administration, after the initial pilot introduction of the measure in 2003. The two administrations of the OAT reading in third grade supports Ohio’s “Fourth Grade Guarantee” that stipulates all students must obtain a “proficient” score on the OAT reading during fourth grade or else face retention or significant intervention efforts (Ohio Legislative Service Commission, 2001).

The OAT measure is a standardized, norm-referenced measure of achievement, aligned with the Ohio Academic Content Standards for a given grade level (Ohio Department of Education, 2009). Raw scores of the OAT are transformed to standardized scores, along with the qualitative descriptions of “basic,” “limited,” proficient,” “accelerated” or “advanced” that describe performance relative to other Ohio same grade level peers. Linear transformation of Rasch ability estimates (theta scores) have been consistently used by ODE since the measure’s introduction, with a standard score of 400 as the cutoff for proficient designation, noting that parallel forms of the OAT exist but are standardized such that the transformed scores can be compared across any subtest or year of administration. For this present study, only OAT subtest scores in reading and mathematics were used. According to Singer and Willet (2003), Rasch ability estimates are ideal data points in longitudinal studies because of their interval and constant metric
properties. Psychometric properties released by the Ohio Department of Education indicate strong structural fit with each subtest and respective Ohio Content Standards according to factor analytic and structural equation modeling techniques (AIR Technical Team, 2008).

**Procedures**

Individual DIBELS ORF and MCA longitudinal results were obtained from the school district’s Child Study Department through a database with the following information: student identification number, grade level, classroom, DIBELS ORF raw scores for fall/winter/spring, and MCA fall/winter/spring raw scores. A matching spreadsheet was provided by the district’s testing and measurement department that contained race, gender, attendance, special education participation status, and OAT reading and math subtest scores for students within this longitudinal cohort. Student names remained anonymous with only unique (non Social Security Number) nominal student identification numbers used for matching purposes across the longitudinal data. Data mining from the district’s mainframe computer system of student records allowed for extraction of any missing data point, if available on district records. Data were reformatted from the spreadsheet format to PASW (Predictive Analytics SoftWare, formerly known as SPSS) Statistics 17.0.3 Graduate Student Version (Prentice Hall, 2009) and HLM for Windows full version 6.08 by Raudenbush, Bryk, and Congdon (2009) as distributed by Scientific Software International.
Data Analysis

A two-level model was developed to analyze the hierarchically structured data of this study. Hierarchically structured models refer to lower (micro-level) data that are nested within higher levels (macro-levels, contexts, or groups). Examples could range from students nested within classes within schools or repeated measurements nested within individual students. In this specific study, students were the micro-level unit containing repeated measurements of the previous year’s DIBELS/MCA raw score with the following year’s OAT reading/math subtest score, which permitted these repeated measures to treated as nested data within individual students. These nested individual students were then combined with each student’s demographic covariates. Two separate models were created, where the first-level linear model for both the OAT reading and OAT math subtest standard scores were regressed onto DIBELS and MCA results, again with these repeated measures representing a nesting within individual student participants. From there a second-level model, in which the regression coefficients of the first-level models were regressed on the second-level explanatory variables was created (Kreft & de Leeuw, 1998). Thus, this study’s data analysis approach represented a two-stage process. In the first stage, the dependent or response variable (OAT reading/math performance at specific time for a specific student) was regressed on level-1 independent or explanatory variable within each unit (DIBELS ORF and MCA) that yielded separate regression equations for each student. In the second stage, coefficient
(i.e., the slope and intercept of the level-1 regression expression) estimates in the first
stage were treated as dependent or response variables while additional explanatory
variables of interest (e.g., race, gender, attendance, and special education participation)
served as additional predictor variables (Pedhazur, 1997). A similar repeated measures
growth HLM methodology gained popularity as used by Sanders (2000) in his seminal
Tennessee Value-Added Assessment System [TVAAS] and Taylor et al.’s (2002)
repeated measurement of reading achievement with various contextual factors, both
discussed in the previous chapter. In addition, Lubienski and Lubienski (2006) employed
hierarchical multilevel modeling to consider differences in school performance by public
and private schools in a similar manner as this current study.

**History of hierarchical linear models [HLM].** Hierarchical or multi-level
models consider the investigation and analysis of datasets where the study observations
are nested within some sort of group. Examples of these nesting phenomena could
include individuals nested within neighborhoods, patients within pharmacological
treatments, or repeated measures of some construct nested within the same person over
time. Indeed, once an appreciation of the nested nature of much social science data is
achieved, the hierarchical organizational frameworks can be observed almost everywhere
(Kreft & de Leeuw, 1998). Across the social sciences, behavioral and most recently,
economic data, have been appreciated to have a hierarchical structure. Within educational
research, these hierarchical structures are gaining prominence, where students are
grouped in classes, classes are grouped in schools, schools in school districts, and so forth. As a result, the researcher is faced with having variables describing individuals, but the individuals are also understood to be grouped into larger or higher-order macro units. In the case of repeated measurement data such as measuring response to psychologic treatment through progressing monitoring instruments, an individual serves as the group, with multiple measurements nested within the individual.

Historically, in traditional quantitative analysis, linear regression models are used for the analysis of such data, with statistical inference based upon the assumptions of linearity, normality, homoscedasticity, and independence. Ideally, only the first of these assumptions should be used yet the aggregation over individual observations may lead to misleading results (Raudenbush & Bryk, 2002). For example, when aggregating student characteristics over larger macro-level classes within a building, all individual information about a specific student is lost in the course of the macro-level analysis. As within-group variation frequently accounts for most of the total variation in the outcome, this loss of information can have an adverse effect on the analysis and lead to distortion of relationships between variables or worse, faulty inference to a larger population. The alternative, disaggregation, implies the assignment of all class, school, and higher-level characteristics to the individual students. Unfortunately, in this process, the assumption of independent observations is violated. Both the aggregation of individual variables to a higher level of observation and the disaggregation of higher order variables to an
individual level have become points of significant methodological concern (e.g. Raudenbush & Bryk, 2002). In addition, serious inferential errors may result from the analysis of complex survey data if it is assumed that the data have been obtained under a simple random sampling scheme when in fact, many such samples are more accidental (e.g. willingness of participants to respond to a survey) in scope due to the hierarchical or contextual similarities across participants.

When considering multi-level or hierarchical organized data, individuals in the same group are likely to be more consistent or similar than individuals in different groups; for example, students in a classroom across a building in comparison to students in a different building’s classroom. Consequently, the variations in results may be due to differences between groups and to individual differences within a group. Variance component models, where variability or scatter may have both a group and an individual component, can be of help in analyzing data of this hierarchical nature. Within these variance models, individual components are independent, yet while the group components are independent within various sub-groups, they are perfectly related within the groups. Random regression models have been developed to model continuous data and also dichotomous repeated measurement data where certain characteristics of the data preclude the use of traditional analysis of variance models (e.g. ANOVA). When using random regression models, however, it is still not easily possible to feature higher order
variables. In order to accommodate both random coefficients and higher order variables, consideration for HLM becomes readily apparent.

HLM analysis allows characteristics of different groups to be included in models of individual behavior. As mentioned previously, most analyses of social science phenomena involves the analysis of data with built-in hierarchies, usually gained through various sampling methodologies. Obviously, this yields a large net for such approaches (Raudenbush, 1988). The formulation of such models and estimation procedures may be seen as an effort to develop a new family of analytical tools that correspond to the classical experimental designs. These models are much more flexible in that they are capable of handling unbalanced or missing data, the analysis of variance-covariance components and the analysis of both continuous and discrete response variables. As the characteristics of individual groups are incorporated into the hierarchical model, the multi-level structure of the data is taken into account and corrected estimates of standard errors are obtained. The exploration of variation between groups, which may be of interest in its own right, depending on research question(s), is also correspondingly facilitated. Validity tests and confidence intervals can also be constructed and stratification variables used in the sample design can be incorporated into the model (Singer & Willet, 2003).

The use of multilevel models has been hampered in the past by extant mathematical models requiring perfectly balanced designs. Because of this, iterative
numerical procedures (e.g. imputation as available in many statistical software programs) must be used to obtain efficient estimates for unbalanced designs. Among the procedures suggested are full maximum likelihood and restricted maximum likelihood estimation (Bryk & Raudenbush, 1992; Pedhazur, 1997). Other approaches include Bayes estimation, the use of Iteratively Reweighted Generalized Least Squares, and a Fisher scoring algorithm, discussed in Appendix C. Increased interest in these models, which are known by various names in the literature, such as hierarchical linear models, multilevel models, mixed-effects models, random-effects models, random coefficient regression models, covariance components models) have led to new developments in this field in recent years and in software programs to handle their analysis.

A separate development in this history was the type of outcome variables considered. While previously interest was confined to continuous outcome variables, statistical theory has been extended and implemented in software such as HLM (Raudenbush, Byrk, & Congdon, 2009) to appropriately handle binary outcomes, ordered categorical outcomes, and multi-category nominal scale outcomes within the hierarchical framework. Indeed, the advances in statistical software for HLM has greatly facilitating the accessibility and incorporation of HLM methods into quantitative research.

Throughout this discussion, an assumption was that each lower-level unit, for example an individual student or patient, was nested within an unique higher-level unit such as a school or treatment. In other words, a one to one relationship was assumed to
define the nesting of units within groups. Excluded from this were hierarchies in which
cross-classification occur, for example where students from multiple neighborhoods may
end up going to multiple schools; a situation where students are "cross-classified" by
neighborhoods and schools. To address this situation and allow for the inclusion of
predictors for more than one "classification" variable where the coefficients of a
individual level model describing the association between individual-level variables and
the outcome for groups defined by the "classification" variables, cross-classified
random-effect models were developed (Raudenbush, 2002). With the arrival of current
HLM software, two-level cross-classified random-effect models are also available.

As interest amplified in multivariate outcome models, such as repeated
measurement data, the inclusion of multivariate models in most of the available
hierarchical linear modeling programs have also correspondingly become available.
These models allow the researcher to study cases where the variance at the lowest level
of the hierarchy can assume a variety of forms/structures. The approach also provides the
researcher with the opportunity to fit latent variable models (Raudenbush & Bryk, 2002),
with the first level of the hierarchy representing associations between fallible, observed
data and latent, "true" data. An application that has received attention in this regard
recently is the analysis of item response models, where an individual’s "ability" or "latent
trait" is based on the probability of a given response as a function of characteristics of
items presented to an individual. Clearly, given this history, delimitations of past methods
such as analyses of variability, and the significant of the ignoring the dependency of nested data, HLM is quickly establishing itself as the de-facto methodology for much social science research (e.g. Grimm & Yarnold, 1995; Kreft & de Leeuw, 1998; Raudenbush & Bryk, 2002).

**Comparison of HLM to previous quantitative statistical analysis techniques.**

Assorted methodological approaches could have been proposed but technical limitations of each precluded their consideration. Prior to the development of multilevel modeling technologies (hierarchical linear modeling [HLM]), multiple regression or analysis of covariance (ANCOVA) may have been an appropriate analytical methodology. ANCOVA could have been employed but would be limited in its utility. Although ANCOVA allows modeling the context, context-specific characteristics could not be modeled due to the hierarchical nature of repeated measures nested within students. This would have reduced the utility of the present study in that the strength of relationship of second-level variables ascertained through the slopes (e.g., attendance, race, gender, special education participation) for DIBELS ORF and MCA would have been lost. Similarly, the amount of overall variability contributed by each second level explanatory variable on the intercept would have precluded this model from predicting the OAT reading/math subtest standard scores. Kreft and de Leeuv (1998) also noted that ANCOVA in school-based research is problematic because it fundamentally assumes experimental, randomized groups where observations are assumed to be independent.
replications of one another, devoid of intra-class correlational [ICC] effects. HLM is also conceived to be superior to multiple regression in school-effects research because of its ability to (a) predict student achievement as a regressed function of student, classroom or school-level characteristics while accounting for the intra-class correlational effects; (b) model contributatory effects of background variables and explain differences in these effects across groups; (c) model between and within-class variability simultaneously; and (d) produce better estimates of predictors of student outcomes by using nested hierarchies (Bryk & Raudenbush, 1992, as cited in Wojcik, 2008).

**HLM in the current study.** For this present investigation, two levels of equations were proposed: Level-one represented a general multiple regression where future OAT reading/math achievement scores were regressed onto individual students’ DIBELS ORF and MCA scores from the previous year; Level-two represented the strength of relationship and contributions among various predictor variables of race, gender, attendance, and special education participation in the partition of variance provided by each to the larger component of OAT reading/math subtest standard scores. DIBELS ORF and MCA scores were paired around the OAT reading/math scores to ensure that the regression coefficients from level-1 adequately reflected the slopes and intercepts of the level-2 equations. In addition, DIBELS/MCA were grand-mean centered while race, gender, attendance, and special education status were grand-mean centered, as further described in Chapter 3.
Restructuring of PASW data files was necessary to align data from a traditional “wide” format to a columnar case style where individual student data points were grouped as cases. In addition, two separate data files were created for each dependent variable of OAT Reading and Math subtest standard scores such that one file contained the paired previous year’s DIBELS/MCA results with the future OAT reading/math standard score while the second file included all the demographic attributes of each student. The student’s unique district identification number served as the constant case group identification variable to link or identify students between the two files in the HLM software.

**Level-One Equations**

\[ Y_{OATR} = \pi_0 + \pi_1 (DIST\_DIBELS) + r_{i-OATR} \]

and

\[ Y_{OATM} = \pi_0 + \pi_1 (DIST\_MCA) + r_{i-OATM} \]

where:

- \( Y_{i-OATR} \) is the predicted OAT reading subtest score for student, \( i \)
- \( Y_{i-OATM} \) is the predicted OAT math subtest score for student, \( i \)
- \( \pi_{0i} \) is intercept or average OATR/OATM score
- \( \pi_{1i} \) is slope or change in OATR/OATM by DIBELS/MCA
DIBELSi is the specific DIBELS score for student, i, grand-mean centered

MCAi is the specific MCA score student, i, grand-mean centered

rif-OATR is residual error on OAT reading for student, i

rif-OATM is residual error on OAT math student, i

Level-2 analysis included student specific variables to examine whether OAT reading and OAT math could be predicted from DIBELS/MCA, once the predictor variables of race, gender, attendance, and special education participation were controlled.

Level-Two Equations

For brevity, only one set of each intercept and slope equations are presented below. However, to be complete, it should be noted that two separate \( \pi \) second-level equations could be written, the first for OAT math and the second for OAT reading, respectfully. In reviewing these equations, the reader is asked to acknowledge both OAT reading and math for each covariate as indicated.

The intercept-specific equation accounts for the predicted OATR/OATM standard score when DIBELS/MCA and individual student predictor variables are controlled:

\[
\pi_{0i} = \gamma_{00} + \gamma_{01}(RACE_i) + \gamma_{02}(GENDER_i) + \gamma_{03}(ATTEND_i) + \gamma_{04}(SPED_i) + r_{01}
\]
The slope-specific equation describes how each individual student attribute moderates the strength of relationship in DIBELS/MCA scores in their contribution to OAT reading/math subtest performance:

\[ \pi_i = \gamma_0 + \gamma_1 (RACE_i) + \gamma_2 (GENDER_i) + \gamma_3 (ATTEND_i) + \gamma_4 (SPED_i) + \epsilon_i \]

where; noting that each level-2 variable is grand-mean centered:

- **RACE\_i** is race of students, coded as Black or White (1=black, 0=white)
- **GENDER\_i** is biologic sex of student, coded as (1=male, 0=female)
- **ATTEND\_i** is number of days of total attendance between 2002 and 2006
- **SPED\_i** is individual student participation in special education and related services programming, coded as Yes or No (0=no, 1=yes)

**Intercept-specific variables.**

- \( \gamma_{00} \) is average score in OAT reading/math performance for students who have average attendance.
- \( \gamma_{01} \) is the average score in OAT reading/math performance between genders, controlling for race, attendance, and special education.
- \( \gamma_{02} \) is the average score in OAT reading/math performance by racial differences after controlling for the effects of gender, attendance, and special education
- \( \gamma_{03} \) is the average score in OAT reading/math performance based
on participation in SPED programming after controlling for race, gender, and attendance

$\gamma_{04}$ is the average score in OAT reading/math performance for higher attending (q.v., attendance above the grand-mean) students after controlling for race, gender and special education

$r_{0i}$ is the residual error for student $i$ on OAT math/reading, controlling for all other variables

**Slope-specific variables.**

$\gamma_{10}$ is the magnitude of relationship of OAT reading/math performance for students who are White, female, high attendance, and no special education intervention

$\gamma_{11}$ is the magnitude of relationship between OAT reading/math performance and gender groups after controlling for race, attendance, and special education

$\gamma_{12}$ is the magnitude of relationship between OAT reading/math performance and race after controlling for gender, attendance, and special education

$\gamma_{13}$ is the magnitude of relationship between OAT reading/math performance based on participation in SPED programming after controlling for race, gender, and attendance
\( \gamma_{14} \) is the magnitude of relationship between OAT reading/math performance and higher attendance (q.v., attendance above the grand-mean) after controlling for race, gender, and special education.

\( r_{1i} \) is the residual error for student \( i \) on OAT math/reading, controlling for all other variables.

This study proposed the creation of a model of variables related to OAT reading and mathematics performance amongst elementary aged students, by using DIBELS ORF and MCA with various demographic variables such as race, gender, attendance, and participation in special education intervention. This study identified first whether DIBELS oral reading fluency could predict OAT reading achievement scores and, separately, whether informal Math Concepts and Applications could predict OAT math performance as the first-level equations and consistent with some extant literature. On the second-level equations, both the OATR/OATM standard score prediction and relationship strength between the various independent variables (e.g., race, gender, attendance, special education programming) were explored. The overall intent of this study was to demonstrate the amount of predictiveness or ability to discriminate each of the above various variables in partitioning OAT math and reading performance. It was hypothesized that DIBELS reading fluency would have a positive predictive relationship in OAT reading performance and MCA would have a positive predictive relationship in OAT math performance. On the second-level equations and models, it was predicted that
there would be significant differences in OAT reading/math performance across race, gender, attendance, and special education intervention both in DIBELS/MCA correspondence (slopes) and in their prediction of OATR/OATM (intercepts) standard scores.
CHAPTER 3: RESULTS

Introduction

The following chapter serves to summarize the results of the present study, which investigated how the response variables of Ohio Achievement Test Reading (OAT-R) and Mathematics (OAT-M) subtests were regressed onto repeated measures of informal assessment, including Dynamic Indicators of Basic Early Literacy Skills Oral Reading Fluency (DIBELS ORF) and Math Concepts and Applications (MCA), respectfully, for elementary school students. In addition, several individual student explanatory covariates including race, gender, attendance, and special education participation were explored to determine how each impacted the strength of relationship of DIBELS ORF and MCA as well as how both of those informal measures predicted OAT subtest standard scores after controlling for each student-level variable. The final outcome product was to develop a system of equations through hierarchical lineal modeling that addressed the above and met the unique data analysis challenges of investigating hierarchically-organized, repeated measures of individual student reading and math performance.

Description of the Sample

Erosion of the original proposed sample. Participants for the current study were all public school students in grades 1-5 at two separate elementary buildings within a large Northern Ohio urban school district, assessed on regularly occurring intervals
between 2002 and 2006. Like many urban settings, this particular school district has struggled with attrition and urban flight (Clotfelter, 2001) of their average daily membership (ADM; q.v., student body) during the past quarter century.

When this study was originally devised, the intended methodology was to employ a longitudinal design where students from both school buildings would have been paired with their DIBELS/MCA and OAT reading (OATR) and OAT mathematics (OATM) subtest results across 5 years, ranging from 2001 to 2005. However, a strikingly decreasing population of students was observed, as shown in Table 4, making such a longitudinal pairing of students untenably confounded by extensive missing data points. For example, a student may have had two years of DIBELS data but no future OATR results or perhaps one OATM result was available without any corresponding MCA informal scores. In addition, because of the initial deployment of the OAT by ODE, starting in third grade, some students did not have the possibility to participate in all OATM subtest administrations, and the number of OATR versus OATM administrations were unbalanced, further limiting the number of subjects. Consequently, had this original longitudinal pairing methodology been used, descriptive results reveal how the original cohort of participants from grade 1 to grade 5, who would have had sufficient DIBELS/MCA scores to pair with subsequent OATR/OATM results, came to exhibit figurative longitudinal respondent mortality in the 5 years of this data collection at 52% for building “A” and 73% at building “B.” Such an occurrence is oftentimes discussed as
a challenge of longitudinal data collection efforts and designs (e.g., Shaughnessy & Zechmeister, 1994).

Table 4

*Originally Proposed Longitudinal Sample Attrition by School Building*

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>“A”</td>
<td>54</td>
<td>43</td>
<td>36</td>
<td>32</td>
<td>26</td>
</tr>
<tr>
<td>“B”</td>
<td>67</td>
<td>53</td>
<td>46</td>
<td>34</td>
<td>18</td>
</tr>
<tr>
<td>Totals</td>
<td>121</td>
<td>96</td>
<td>82</td>
<td>66</td>
<td>44</td>
</tr>
</tbody>
</table>

In understanding the original methodology’s attrition rate from \( n = 121 \) to \( n = 44 \), those diminishing students either willingly relocated to another building, were unilaterally transferred to another building due to disciplinary sanction, or else withdrew from the district either for a specific timeframe or altogether. In addition, building “B” was scheduled for closure at the conclusion of the 2005-06 academic year, which may have hastened some students’ departures in hopes of establishing themselves at another
building and high school cluster of schools, as the district is geographically organized, before compulsory closure dictated the same. Such attendance erosion trends appear to be commonplace within this district as staff anecdotally note the high rate of pupil itinerancy.

**Development of a new sampling methodology.** As a consequence of this high attrition rate, and to ensure adequate sample size without extreme missing data contamination, the entire universe of students, representing *all* students who attended either of these two buildings *at some point* in their elementary career was examined, \( N = 1455 \). From this larger cohort, a sampling approach was used where the previous year’s spring informal raw scores on DIBELS/MCA were paired with the subsequent year’s OATR/OATM standard score, now restricted to academic years 2002 to 2006, to ensure a one to one matching pair of each data set. This methodology was described in Chapter 2 for reference and allowed for a considerably more robust sample size devoid of missing or awkwardly paired data. Further discussion of missing data and its implications for HLM is provided in Appendix C for the interested reader.

Composition of this new sample, \( n_{\text{total}} = 162 \), revealed the following demographic features, shown in Table 5 for the sample used throughout the current study, \( n_{\text{reading}} = 71 \) and \( n_{\text{math}} = 91 \), respectfully, cross-tabulated by reading and mathematics content area assessment for ease of visual comparison. Table 6 summarizes total student attendance between 2002 and 2006 for each content area while descriptive statistics for response variables, OAT reading and OAT math, are provided in Tables 7 and 8. Finally, Table 9
provides descriptive statistics on the explanatory variable of DIBELS ORF and Table 10 does likewise for the explanatory variable of MCA.

Table 5

*Descriptive Composition of Sample by Assessment Content Area*

<table>
<thead>
<tr>
<th>Gender</th>
<th>Reading</th>
<th>Math</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>35</td>
<td>46</td>
<td>81</td>
</tr>
<tr>
<td>Female</td>
<td>36</td>
<td>45</td>
<td>81</td>
</tr>
<tr>
<td>Total</td>
<td>71</td>
<td>91</td>
<td>162</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Race</th>
<th>Reading</th>
<th>Math</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>33</td>
<td>26</td>
<td>59</td>
</tr>
<tr>
<td>Non-White</td>
<td>38</td>
<td>65</td>
<td>103</td>
</tr>
<tr>
<td>Total</td>
<td>71</td>
<td>91</td>
<td>162</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Special Education Recipient&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Reading</th>
<th>Math</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>6</td>
<td>9</td>
<td>15</td>
</tr>
<tr>
<td>No</td>
<td>65</td>
<td>82</td>
<td>147</td>
</tr>
<tr>
<td>Total</td>
<td>71</td>
<td>91</td>
<td>162</td>
</tr>
</tbody>
</table>

Note: <sup>a</sup> No distinction of disability condition was made other than binomially designating whether a student had been identified as receiving special education. The exact point of identification and nature of the disability are unknown across the present sample.
Table 6

Descriptive Statistics of Sample’s Attendance by Academic Content Area

<table>
<thead>
<tr>
<th></th>
<th>Reading</th>
<th>Mathematics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total Days of Attendance between 2002-2006</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>M</strong></td>
<td>668.48</td>
<td>665.19</td>
</tr>
<tr>
<td><strong>SD</strong></td>
<td>23.51</td>
<td>33.14</td>
</tr>
<tr>
<td><strong>Minimum</strong></td>
<td>545</td>
<td>446</td>
</tr>
<tr>
<td><strong>Maximum</strong></td>
<td>692</td>
<td>692</td>
</tr>
<tr>
<td><strong>Kurtosis</strong></td>
<td>10.14</td>
<td>22.40</td>
</tr>
<tr>
<td><strong>Skew</strong></td>
<td>-2.53</td>
<td>-4.05</td>
</tr>
<tr>
<td><strong>n</strong></td>
<td>71</td>
<td>91</td>
</tr>
</tbody>
</table>

*Note.* The typical school year calendar has 180 days scheduled but varies yearly due to the presence of calamity days and other unexpected events. Assuming no calamity days or other reasons for lost instructional time, such as student absenteeism due to illness, the maximum possible attendance was 720 days per student during the 4 year period of the study.
Table 7

*Descriptive Statistics for OAT Reading Subtest Standard Scores*

*by Grade, n = 71*

<table>
<thead>
<tr>
<th></th>
<th>Third</th>
<th>Fourth</th>
<th>Fifth</th>
<th>Sixth</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>422.56</td>
<td>425.07</td>
<td>423.72</td>
<td>424.38</td>
</tr>
<tr>
<td>SD</td>
<td>21.66</td>
<td>27.76</td>
<td>30.21</td>
<td>25.30</td>
</tr>
<tr>
<td>Minimum</td>
<td>376</td>
<td>366</td>
<td>350</td>
<td>358</td>
</tr>
<tr>
<td>Maximum</td>
<td>503</td>
<td>486</td>
<td>500</td>
<td>487</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>2.12</td>
<td>-.30</td>
<td>.30</td>
<td>.21</td>
</tr>
<tr>
<td>Skew</td>
<td>.72</td>
<td>.13</td>
<td>.27</td>
<td>-.02</td>
</tr>
</tbody>
</table>

*Note.*  Most students had the opportunity to take the OAT Reading subtest twice during third grade as ODE began deploying the measure. To ensure consistency of comparison across the previous year’s spring DIBELS to follow year’s OATR, only the spring OATR was used for third grade.

*Proficient status on OAT requires a transformed standard score ≥400*
Table 8

Descriptive Statistics for OAT Mathematics Subtest Standard Scores
by Grade, n = 91

<table>
<thead>
<tr>
<th></th>
<th>Fifth</th>
<th>Sixth</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M^a$</td>
<td>398.62</td>
<td>406.36</td>
</tr>
<tr>
<td>$SD$</td>
<td>24.30</td>
<td>33.21</td>
</tr>
<tr>
<td>Minimum</td>
<td>349</td>
<td>303</td>
</tr>
<tr>
<td>Maximum</td>
<td>463</td>
<td>479</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>-.21</td>
<td>.51</td>
</tr>
<tr>
<td>Skew</td>
<td>.20</td>
<td>-.18</td>
</tr>
</tbody>
</table>

$^a$ Proficient status on OAT Mathematics requires a transformed standard score $\geq 400$. 
Table 9

*Descriptive Statistics for Spring DIBELS Oral Reading Fluency by Grade, n = 71*

<table>
<thead>
<tr>
<th>Grade</th>
<th>M</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
<th>Kurtosis</th>
<th>Skew</th>
</tr>
</thead>
<tbody>
<tr>
<td>Second</td>
<td>106.85</td>
<td>37.06</td>
<td>33</td>
<td>240</td>
<td>1.59</td>
<td>.58</td>
</tr>
<tr>
<td>Third</td>
<td>117.45</td>
<td>29.56</td>
<td>43</td>
<td>180</td>
<td>-.05</td>
<td>-.20</td>
</tr>
<tr>
<td>Fourth</td>
<td>129.23</td>
<td>36.90</td>
<td>49</td>
<td>214</td>
<td>-.00</td>
<td>.11</td>
</tr>
<tr>
<td>Fifth</td>
<td>132.10</td>
<td>31.37</td>
<td>44</td>
<td>202</td>
<td>.48</td>
<td>-.29</td>
</tr>
</tbody>
</table>

Table 10

*Descriptive Statistics for Spring Math Concepts and Applications by Grade, n = 91*

<table>
<thead>
<tr>
<th>Grade</th>
<th>M</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
<th>Kurtosis</th>
<th>Skew</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fourth</td>
<td>20.35</td>
<td>7.01</td>
<td>6</td>
<td>40</td>
<td>-.25</td>
<td>.24</td>
</tr>
<tr>
<td>Fifth</td>
<td>12.48</td>
<td>6.29</td>
<td>0</td>
<td>30</td>
<td>.51</td>
<td>-.18</td>
</tr>
</tbody>
</table>
Sample and Data Requirements for HLM

All statistical procedures have underlying assumptions, which, depending upon the degree to which they are fulfilled or violated, may change substantive research conclusions. Employment of hierarchical linear modeling methodology presupposes several assumptions related to the data and sample characteristics that need to be considered prior to the development of these models (Raudenbush & Bryk, 2002). First, is the assumption of linearity, meaning that the function forms are linear and can be graphed with reasonable concurrence to a line equation form that is otherwise not quadratic. Second, the normality of the scores and residuals such that level-1 residuals are normally distributed and level-2 random effects ($r_i$) have a multivariate normal distribution. Third is homoscedasticity, which refers to the residual variance at level-1 remaining constant. Fourth, independence of residuals in that the level-1 and level-2 residuals are not confounded by intercorrelation effects. Finally, independence, noting that observations at the highest level, are independent of one another. These slightly differ from ordinary least squares (OLS) assumptions of regression that emphasize linearity, normality, homoscedasticity, and independence (Snijders & Bosker, 1999). The following section will investigate these assumptions in the current study.

Linearity

HLM models assume that the growth trajectory or regression function for each datum, when computed collectively, is linear. Examination of this assumption requires
the use of scatter plots or the measurement of a linearity statistic such as curve fitting with R-squared difference tests or eta \( \eta \), the coefficient of nonlinear correlation. In the present study, scatter plots of the DIBELS ORF and MCA, as administered over time, were plotted. Overall, no visual evidence of non-linear trends was observed suggesting that the assumption of linearity had not been violated.

**Normality**

Using descriptive techniques for the response variables of OAT reading and OAT mathematics, as reported in Tables 5 and 6, no significant evidence that the assumption had been violated was noted across all administrations of these measures. Specifically, the mean and standard deviation results were consistent with psychometric data provided by the Ohio Department of Education (AIR Technical Team, 2008). Measures of skewness and kurtosis reveal generally whether data are normally distributed with a bell-shaped curve, with the majority of observation clustered around the mean. Skewness indicates whether the frequent distribution’s graph is centered in the middle of the distribution with acceptable measures of ±1. All OAT subtests reflected a minimal skewness measure between ±1, suggesting the vast majority of scores clustered around the mean. Kurtosis measures reflect the extent to which data are concentrated in the peak versus the tail with acceptable measures also ±1. Five of six OAT measures reflected strong clustering around the mean except for the OAT reading third grade test, which revealed some moderate kurtotic tendencies. Such a finding may not violate the
assumption of normality and is consistent with expected findings since those taking the spring OAT third grade reading test would represent those students’ first exposure to the OAT measure and such a high-stakes experience. Histogram presentation for all six OAT subtests yielded visual data of normal distribution, particularly when a normal curve was overlaid, although a slight negative skew was noted in composite OAT reading, likely due to the third grade issue above, versus composite OAT math, which was highly bell-shaped across the two administration points.

**Multicollinearity**

The concept of collinearity, or multicollinearity when referring to collinear relations between more than two variables, describes the potential adverse effects of correlated independent variables on the estimation of regression statistics. In the simplest sense, collinearity describes variables that are nearly perfectly correlated, which can have disastrous implications on regression statistics that could render them useless according to some literature (e.g., Pedhazur, 1997). In the most severe instance of this situation, regression coefficients could have imprecise estimates whereby even slight fluctuations in data (e.g., measurement error, random error) could lead to substantial fluctuations in the sizes of regression coefficients and the qualitative aspects of their signs.

Examination of (multi)collinear relationships can be accomplished through the use of correlational computation between various predictor variables, which in the present study included DIBELS, MCA, gender, race, attendance, and special education
status. Low, non-statistically significant correlations amongst these variables would likely indicate a minimal or non-existent condition of multicollinearity for variables included in the model.

**Examination of mathematics multicollinearity.** In mathematics, the level-2 covariates, including race, gender, attendance, and special education affiliation, exhibited no evidence of any significant correlation ($p > .05$) between one another. General examination of correlation coefficients was also visually unremarkable for any notable relationship, with Pearson $r$ values ranging from -.147 to 0.03. Not surprisingly, across the informal MCA measurements in fourth and fifth grade, moderately significant correlations were noted between fourth grade MCA and fifth grade MCA, $r = 0.50$, $p < .01$, which should be anticipated if these measures exhibit high predictive validity across each year’s form, general construct validity, and test-retest reliability across increasing years of administration. Likewise, MCA fourth grade was modestly and inversely related to special education status ($r = -.314$, $p < .01$) and MCA fifth grade with special education exhibited the same inverse trend, $r = -.420$, $p < .01$. The later findings of modest to moderate relationship between MCA and special education status are equally expected as those informal measures assess mathematical skills, which may be presumed to be weaker across students receiving special education and related services.

**Examination of reading multicollinearity.** Across reading level-2 covariates, again including race, gender, attendance, and special education participation, no evidence
of statistically or practical significance were noted. General Pearson \( r \) values ranged from -.15 to .00, revealing an absence of any relationship. Similar to MCA, DIBELS measures revealed high relationship across successive years of administration, \( p < .01 \), with \( r \) values ranging from .782 to .871 as would be expected given the well-established psychometric properties, including test-retest reliability and construct validity reported in the literature (e.g., Fuchs, Fuchs & Hosp, 2001). Also similar to MCA and special education affiliation, modest inverse relationships were noted across the DIBELS raw scores and special education participation (\( r \) values ranged from -.266 to -.344), all of which were statistically significant, \( p < .01 \). The same logic can be applied that those students identified for special education services may exhibit difficulty in reading, promoting their identification as students with disabilities requiring special education and related services.

**Options for redressing multicollinearity.** Correcting for the presence of multicollinearity across yearly DIBELS and MCA administrations as well as both informal measures inverse relationships to special education identification could be remedied through several approaches. First, leaving the data that will create the model as is, noting that it may not affect the conditional fitted model because the predictor variables likely follow the same pattern of multicollinearity as the data upon which the regression model is based. Second, mean-center these predictor variables when constructing the models, which mathematically would not affect the regression but serves
to overcome problems of estimation and rounding as present in software algorithms, not
to mention facilitate interpretation of results. Other more austere methods could include
dropping one of the explanatory variables, compute a Sharpley value (e.g., Lipovestky &
Conklin, 2001) and certainly to obtain more data by increasing sample size.

Depending on perspective, multicollinearity may not be the foremost threat to
HLM in that it does not affect the reliability of the forecast, instead influencing the
interpretation of the explanatory variables and the estimate results. Indeed, as long as the
collinear relationships in the independent variables remain stable over time, the
prediction is likely to be less influenced (Lubienski & Lubienski, 2006), which appeared
to be the instance within this study.

**Distribution of Errors and Homoscedasticity**

Across hierarchical linear modeling techniques, homoscedasticity refers to the
assumption that level-1 residual variance remains constant with a mean of 0 or what is
often called the homogeneity of variance. In addition, independence of residuals is
assumed across both level-1 and level-2 where the residuals are not intended to be
correlationally significant in relationship. Finally, the residuals across both levels of the
models are assumed to be normally distributed. Completion of the HLM building process
allows computational software such as HLM 6.08 (Raudenbush, Byrk, & Congdon, 2009)
to statistically calculate these data.
According to Raudenbush and Byrk (2002), heterogeneity of levels 1 and 2 variances can have several causes: (a) one or more level-1 predictor variables may have been omitted from the model that may have unequal variance across groups; (b) the effects of a level-1 predictor that is random or non-randomly varying may have been treated as fixed or omitted; (c) one or more units have bad data; and (d) non-normal data with heavy tails or kurtosis in their distributional shape may be present. The most commonly used test statistic in HLM for homogeneity of variance is the $H$ statistic, generally considered an appropriate test when the data are normal and sample sizes per unit $\geq 10$, denoted by the following equation:

$$H = \sum d_j^2$$

where there is a large sample $\chi^2$ distribution with $J-1$ degrees of freedom under the homogeneity hypothesis.

In both reading and mathematics, level-1 heteroscedasticity of errors was noted as computed by the $H$ statistic and the chi square computation revealing significance: $Y_{OATM} = \chi^2 (90) = 153.93, p < .01$ and $Y_{OATR} = \chi^2 (70) = 152.81, p < .01$. Although this violation represents a possible study limitation due to impinging upon a basic assumption of HLM, heteroscedasticity, per se, is not considered a serious problem for estimation of level-2 coefficients or their standard errors, which was the foremost rationale for the present study to determine those level-2 coefficients. The $H$ test is very sensitive to even minute
violations in the normality of the data, which certainly was present in the OATR third grade assessment. In addition, extreme individual values of \( d_j \) can produce serious increases in the magnitude of \( H \), which can be readily observed through the wide range exhibited across all level-1 variables, including the response variables, OATR and OATM, as well as the explanatory variables of DIBELS and MCA. Further investigation using scatter plots of the residual errors produced by both equations noted that \( H \) was likely inflated due to the large range across scores, which is not easily overcome in real-life practice as classrooms are not perfectly or randomly selected in student composition and considerable variability exists in the measurement of achievement characteristics when students are heterogeneously grouped as the present study’s participants were.

**HLM Model Building**

**Towards a Process of Hierarchical Linear Model Building**

The process of hierarchical linear model development employs the usage of empty or unconditional models upon which subsequent predictor variables are added to determine the amount of variance partitioned out by each subsequent addition and whether such an addition of a covariate or explanatory variable adds or reduces composite variability. In the present study, two preliminary empty models were developed, the full unconditional models for reading and mathematics, respectfully. From there, additional covariates were added to each unconditional or empty model, beginning
with DIBELS and MCA, then culminating with the individual student covariates of race, gender, attendance, and special education, through a successive process of introducing each variable and calculating whether variability has changed. The final models for each response variable yield the overall prediction model that can be used to accept or reject the study’s research hypotheses by noting how variance is partitioned through the inclusion of subsequent explanatory variables. This final model is called the full conditional model.

The Unconditional Model is oftentimes called the “empty” or “null” model, or the familiar one-way ANOVA with random effects from elementary statistics, because it only contains the response variable. Generically, unconditional models are denoted by the following first and second level equations:

Level-1: \( Y_{ij} = \beta_{oi} + r_{ti} \) (denoting persons nested within groups)

Level-2: \( \beta_{ij} = r_{ti} \)

or

Level-1: \( Y_{ij} = \pi_{oi} + r_{i} \) (denoting measures nested within persons)

Level-2: \( \pi_{oi} = r_{i} \)

In the case of the present study, two separate unconditional model regression equations can be offered, noting the change for each response variable of OATR/OATM and inclusion of \( \pi \) for \( \beta \) coefficients. This latter Greek letter substitution signifies that the
individual measures are nested data containing repeated measures within individuals
instead of individuals nested within some group or context:

Level-1: \( \text{OATR}_{it} = \pi_{oi} + r_i \)

Level-2: \( \pi_{oi} = r_i \)

and

Level-1: \( \text{OATM}_{it} = \pi_{oi} + r_i \)

Level-2: \( \pi_{oi} = r_i \)

Both equations simply refer to the whether there are differences in mean achievement of
the OAT reading or mathematics subtest within (level-1) and between (level-2) individual
students (\( i \)).

The second modeled equation is termed the conditional model in the present study
due to the presence of repeated measurements of DIBELS and MCA across students (\( i \)) at
time (\( t \)), that were paired or nested within the repeated measures of OATM/OATR. These
equations can be represented as follows, noting again the difference due to the inclusion
of either OAT reading or OAT math for student (\( i \)), time (\( t \)), MCA and DIBELS
grand-mean centered, and the absence of any additional Level-2 predictor variables:
Level-1 Reading: \[ Y_{\text{OATR}i} = \pi_0 + \pi_{1i} (\text{DIBELS}_i) + r_{i\text{-OATR}} \]

Level-2 Reading: 
\[ \pi_{0i} = \beta_{00} + r_{0i} \]
\[ \pi_{1i} = \beta_{10} + r_{1i} \]

and

Level-1 Math: \[ Y_{\text{OATM}i} = \pi_0 + \pi_{1i} (\text{MCA}_i) + r_{i\text{-OATM}} \]

Level-2 Math: 
\[ \pi_{0i} = \beta_{00} + r_{0i} \]
\[ \pi_{1i} = \beta_{10} + r_{1i} \]

The third and final equation represents the inclusion of all predictor variables, race, gender, attendance, and special education affiliation and permits further specificity through a full slope and intercept specific model.

**HLM Intercept, Slope, Moderator Variables, and Statistical Power**

**Intercept.** The intercept specific equation accounts for the OAT math or reading standard score prediction between various individual \((i)\) student covariates and DIBELS/MCA with their overall partition of variance for OAT reading and mathematics achievement prediction. Like traditional OLS regression, on the base level, an outcome variable is predicted as a function of a linear combination of one or more level-1 variables, plus an intercept (Gelman & Hill, 2007). In current study, this intercept represents the predicted OAT standard score when it is regressed onto DIBELS, MCA as well as the various level-2 student demographic variables. In addition, within the present
analysis, the intercept was left as random as the starting point for the model building sequence and subsequently compared to the level-2 output. As will be discussed within the MCA model building, in cases where there is no random effect and no intercept, then the parameter is not needed in the model. Such a result was found with MCA as a level-1 variable failing to partition OAT-M variability even though the level-2 student demographic variables did partition OAT-M performance.

**Slope.** The slope serves to describe how each individual ($i$) student covariate predicts the strength of relationship in DIBELS/MCA and their respective partitioning of the variance in OAT math/reading performance. That is, how does the influence of various level-1 and level-2 variables influence the relationship between the dependent variables and the various independent variables incorporated at each level? Intercept and slope analysis is also important in HLM as the computed intercepts and slopes from level-1 become dependent variables for the level-2 analysis in the prediction of the outcome variable. This latter example is called intercepts and slopes as outcomes by Raudenbush and Bryk (2002) because it allows the dependent variable, OAT in this study, to be predicted from all other independent variables and cross-level interactions. Each slope from the level-1 equation can be graphed using HLM 6.08 and would reveal the following figures, representing OAT-R and OAT-M, respectfully:
Figure 1: Level-1 OAT-M and Slopes of Grand Mean Centered MCA

Figure 2: Level-1 OAT-R and Slopes of Grand Mean Centered DIBELS
Beyond these level-1 graphs of the slopes, which represent the various regressions, further specificity can be added by manipulating the z-focus feature in HLM 6.08 to include various level-2 covariates of interest.

**Moderator variables.** Classically, a moderator variable is a third variable that determines the magnitude of impact on the causal variable on some outcome (Raudenbush & Bryk, 2002). For example, suppose one is investigating the relationship between the availability of computers and academic achievement. A possible moderator variable could be the use of computers; that is to say, the availability of computers has an effect on academic achievement only to the extent that computers are effectively used in instruction. Within the present study, OAT subtest performance was regressed onto informal measures of reading and mathematics. Moderator variables would include the various level-2 student demographic variables such as race, gender, total attendance and special education participation. In interpreting the relationship between the informal measure and the OAT subtest performance, the presence of those specific student demographic features would modulate or affect the nature of the level-1 relationship.

**Statistical power and sample size.** Statistical power refers inverse of the level or probability of making a type II error, 1-\(\beta\). Such an error involves a failure to reject a null research hypothesis when it is in fact not true (Shaughnessy & Zechmeister, 1994). Usually, the researcher accepts or predetermines the power level as 0.80, meaning that the researcher has an 80% chance of not making a Type II error.
In statistical power, the effect size or the strength of association is the strength of the relationship between two variables of interest. As an effect size becomes greater, the greater its statistical power, and ultimately higher likelihood that the statistical test employed is yielding valid inferences. A greater effect size emphasizes greater statistical power. Sensitivity is referred to the number of true positives out of the total of true positives and false negatives. Generically speaking, sensitivity recognizes the truly “correct” data. This means that a high sensitivity will yield accurate data and therefore a high statistical power, or having less number of Type II errors.

Careful study design ensures that the value of statistical power is appropriately high by considering the size of the sample necessary to satisfy power demands. In most situations, a researcher does not have access to an entire statistical population of interest due to practical or participant-specific factors, which can delimit the number of participants. Yet, as a researcher’s sample size increases, the greater the statistical power and ability to make inferences from the study’s results occur (Spiegel, 2000). While the various computational processes for determining the actual sample size far exceeds this present section, interested readers are referred to Lenth (2001) for additional guidance on selection of adequate sample size prior to onset of a research investigation. Generally, statistical power is increased under the presence of additional observations, whether adding more subjects or further measurements of some attribute of those subjects. Many
commercial software packages such as PASW feature the ability to determine sample size based upon selection of statistical analysis used.

**The whole model.** In building the full model, individual student covariates can also be held constant to examine what happens to the regression model when each are simply added with respect to their influence of relational strength and predictiveness to OAT reading and mathematics standard scores consistent with traditional inferential statistics modeling. This fully conditional model equation can be represented as follows, noting that 4 total equations could be provided, representing OAT reading and OAT mathematics combined with the previous conditional model’s level-1 equations that remain the same:

**Level-2 Intercept:**

\[ \pi_{0i} = \gamma_0 + \gamma_{01} (RACE_i) + \gamma_{02} (GENDER_i) + \gamma_{03} (ATTEND_i) + \gamma_{04} (SPED_i) + r_{0i} \]

**Level-2 Slope:**

\[ \pi_{1i} = \gamma_{10} + \gamma_{11} (RACE_i) + \gamma_{12} (GENDER_i) + \gamma_{13} (ATTEND_i) + \gamma_{14} (SPED_i) + r_{1i} \]

In the presence of algebraic manipulation, a combined full mixed conditional model describing the relationship between OAT reading/mathematics, DIBELS, MCA, race, gender, attendance, and special education affiliation could be written as follows:
OAT = \beta_0 + \beta_{01}(GENDER - \bar{GENDER}) + \beta_{02}(RACE - \bar{RACE}) + \\
\beta_{03}(SPED - \bar{SPED}) + \beta_{04}(ATTEND - \bar{ATTEND}) + \beta_{1c}(DIBEL_{ti} or MCA_{ti} - \bar{DIBEL}_{ti} or \bar{MCA}_{ti}) + \\
\beta_{11}(GENDER - \bar{GENDER}) \cdot (DIBEL_{ti} or MCA_{ti} - \bar{DIBEL}_{ti} or \bar{MCA}_{ti}) + \\
\beta_{12}(RACE - \bar{RACE}) \cdot (DIBEL_{ti} or MCA_{ti} - \bar{DIBEL}_{ti} or \bar{MCA}_{ti}) + \\
\beta_{13}(SPED - \bar{SPED}) \cdot (DIBEL_{ti} or MCA_{ti} - \bar{DIBEL}_{ti} or \bar{MCA}_{ti}) + r_{01} + \epsilon_{ti}

The notion of $\bar{\text{ATTEND}}$, or the double bars overtop any variable in the equation, acknowledges that the variable has been grand mean centered. Centering of variables provide a cogent metric for interpretation (average difference across subjects or conditions) and reduces values from falling outside of possible ranges. Further elaboration of centering methods will be discussed in the section covering the unconditional model development. In addition, while the above mixed equation may appear daunting, the reader should note that it is simply an expanded version of the basic least-squares regression line from elementary statistics, inclusive of the expected intercept($a_0$) and slope ($a_1$) (Spiegel, 2000):

$$Y = a_0 + a_1X$$

In the presence of this full HLM conditional mixed model equation, the research hypotheses can be tested through inferential statistics and subsequent prediction, at least with respect to the current data set and sample, can be proffered. The remainder of this
manuscript will elaborate on building these three specific models then address each research hypothesis in sequential fashion in culmination with Chapter 4, Discussion.

**Unconditional Models: OAT Reading and OAT Mathematics**

Within the unconditional model, or one-way ANOVA with random effects, only one fixed effect is available, $\pi_{00}$, which estimates the mean of OAT reading and mathematics achievement across occasions and individuals, respectfully, by each response variable and unique level-1 equation. These data are presented in Tables 11 and 12 for the response variables OATR and OATM:
### Table 11

*HLM Results for Unconditional Model: OAT Math*

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Coefficient</th>
<th>Robust SE</th>
<th>t-ratio</th>
<th>df</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRCPT2 ($\pi_{00}$)</td>
<td>402.49</td>
<td>2.78</td>
<td>144.58</td>
<td>90</td>
<td>0.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>Variance Component</th>
<th>df</th>
<th>$\chi^2$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between person, $r_{0t}$</td>
<td>566.03</td>
<td>90</td>
<td>436.35</td>
<td>0.00</td>
</tr>
<tr>
<td>Within person, $e_{it}$</td>
<td>294.17</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Deviance = 1687.93 with 2 parameters
Table 12

*HLM Results for Unconditional Model: OAT Reading*

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Coefficient</th>
<th>Robust SE</th>
<th>t-ratio</th>
<th>df</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRCPT2 ($\pi_{00}$)</td>
<td>423.93</td>
<td>2.80</td>
<td>151.53</td>
<td>70</td>
<td>0.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>Variance Component</th>
<th>df</th>
<th>$\chi^2$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between person, $r_{0t}$</td>
<td>519.25</td>
<td>70</td>
<td>888.03</td>
<td>0.00</td>
</tr>
<tr>
<td>Within person, $\epsilon_{it}$</td>
<td>177.73</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Deviance = 2450.79 with 2 parameters

**Unconditional model for OAT reading.** In reading, the grand mean of the OAT reading subtest was estimated to be 423.93 ($SE = 2.80$) across all measurements of the OAT reading in this sample. The $t$-ratio was statistically significant, $t(70) = 150.46, p < .01$, mainly meaning that the true grand mean of reading achievement of all students who took the OAT reading between grades 3 and 6 is non-zero. The $\chi^2 (70, n = 71) = 888.03, p < .01$ for between-person variance indicates that there is significant unexplained variability between individual students and their mean OAT reading achievement. A
computation of the intraclass correlation coefficient [ICC], \( \rho \), can be computed to describe the proportion of total variance in OAT reading achievement that exists across individual OATR measurements as follows:

\[
\rho = \frac{\tau_{00}}{(\tau_{00} + \sigma^2)} = \frac{519.25}{(519.25 + 177.73)} = 0.745
\]

This suggests that 74.5\% of the variance in OAT reading achievement is present between individual students while 25.5\% (100 * (1-0.745)) is present within individual students. Computation of \( \rho \) is useful in HLM to determine the appropriateness of continuing with model building as when \( \rho \) approaches 0 or is negative, HLM is contraindicated. With the current value of \( \rho = 0.745 \), there is further support for continuing the HLM building process. Moreover, while a high ICC may be usually considered an assumptional violation, in this instance, such a result should be present as more variability should exist across students than within individual students.

**Unconditional model for OAT mathematics.** For OAT mathematics subtests, the grand mean of the OAT mathematics subtest was estimated to be 402.49 (SE = 2.78), which is remarkably consistent with the OAT mathematics mean of 400 as provided by ODE (AIR Technical Team, 2008). A statistically significant \( t \)-ratio was also observed, \( t(90) = 143.78, p < .01 \), similarly indicating that the grand mean is also non-zero. Calculation of \( \chi^2(90, n = 91) = 436.35, p < .01 \), suggests that there is considerable unexplained variability between individual students in mean OAT math achievement.
Computation of the ICC [$\rho$] reveals that 65.8% of individual math achievement variability exists *between* individual students as follows:

$$\rho = \frac{\tau_{00}}{(\tau_{00} + \sigma^2)} = \frac{566.03}{(566.03 + 294.17)} = 0.658$$

Separately, 34.2% of the variability occurs *within* individual students by calculating, 1- $\rho$.

Similar to reading, because $\rho = .658$, there is further support to continue the HLM process. Akin to the reading results, such a finding should be present, as a higher amount of variability should be expected *between* individuals than *within* each student.

**Conditional Models for OAT Reading & Math**

In this next model production phase, the addition of the explanatory variables, \(DIBELS_t\) and \(MCA_i\), allowed for computation of two fixed effects, $\pi_{00}$ and $\pi_{10}$. These incorporated both the OAT reading and mathematics subtest achievement along with the presence of multiple measurements of the DIBELS and MCA, respectfully, paired by the previous year’s informal measurement with the subsequent year’s OAT results. By adding this next level-1 variable to the original empty unconditional model, additional variance is intended to be explained by the measurement of these informal measures and how they can contribute to the prediction of relational strength and OAT achievement standard scores. These results are provided in Tables 13 and 14:
Table 13

*HLM Results for Conditional Model: OAT Math*

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Coefficient</th>
<th>Robust SE</th>
<th>t-ratio</th>
<th>df</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRCPT2 ($\pi_{00}$)</td>
<td>402.49</td>
<td>2.78</td>
<td>144.58</td>
<td>90</td>
<td>0.00</td>
</tr>
<tr>
<td>Relationship ($\pi_{10}$)</td>
<td>-0.82</td>
<td>0.22</td>
<td>-3.81</td>
<td>180</td>
<td>0.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>Variance Component</th>
<th>df</th>
<th>$\chi^2$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between person, $r_{0t}$</td>
<td>552.56</td>
<td>90</td>
<td>418.93</td>
<td>0.00</td>
</tr>
<tr>
<td>Within person, $e_{it}$</td>
<td>301.46</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Deviance = 1690.93.01 with 2 parameters
Variable centering. The development of the second model, the conditional model, for each OAT subtest, requires a decision to be made about whether to center the predictor variables, DIBELS_{it} and MCA_{it}.

Leaving a predictor variable un-centered, which is also referred to as using the natural metric for that variable, is only recommended if the value for the predictor variable can be meaningfully equal to 0. An example of the latter could occur in a binominal or dichotomous variable such as gender, which is traditionally coded 0 or 1.
When 0 is not a meaningful value for a predictor variable, the estimate of the intercept will be likely arbitrary, divergent from the metric of the variable, and could be estimated with poor precision. This represents a salient issue and interpretative quandary in HLM that could compromise the entire model’s understandability. When selecting between centering options, once un-centered is eschewed, the researcher is left to decide between group mean and grand mean centering (Snijders & Bosker, 1999).

Grand mean centering addresses the problems of intercept estimation in the original metric because 0 will fall in the middle of the distribution of predictors. This allows the intercept estimates to have more precision and leads to them being easily interpretable by allowing a representation of the grand-mean value for a person with a (grand) mean average on every predictor. In other words, the results are centered around the overall average of everyone in the study; however, interpretation of the slope does not change. Group mean centering allows predictors to be organized around the mean value for the cohort or group they belong. This allows the intercept to be interpreted as the average outcome for each group; in other words, centered about one’s own level-1 “group” average (Raudenbush & Bryk, 2002). Because the unit of interest in the present study is interpretation of average effects across each paired measurement of DIBELS/MCA with their corresponding OAT subtest scores and the intercept needs to be interpretable in such a manner, grand mean centering was applied to MCA and DIBELS as well as the predictor variables all the level-2 variables. Generically speaking, in HLM,
dummy variables are left un-centered, continuous variables at level-1 are group mean centered, and continuous variables at level-2 are grand-mean centered (Boyle & Willms, 2001; Gelman & Hill, 2007).

By grand mean centering the independent variables of MCA and DIBELS, these scores were able to be interpreted as the overall effect for each measurement on the natural metric of the OAT standard score. These centering applications were notated by double bar over the level-1 explanatory variables in the conditional equations, e.g., $\overline{MCA}$ and $\overline{DIBELS}$, and with double bars in subsequent level-2 variables, e.g., $\overline{ATTEND}$, $\overline{SPED}$, $\overline{GENDER}$, and $\overline{RACE}$.

**Conditional model for OAT mathematics.** Across the conditional model for OAT mathematics, two fixed effects were estimated in this subsequent model, $\pi_{00}$ and $\pi_{10}$, representing the average OAT math achievement (402.49, $SE = 2.76$) and change in math achievement for every point MCA changes (0.05, $SE = 0.22$) on the metric of OAT standard score points per year. The fixed effect of INTRCEPT1($\pi_{00}$) was statistically significant, $t(90) = 145.75, p < 0.01$, while MCA SLOPE ($\pi_{10}$), $t(180) = 0.23, p = 0.82$ was unsignificant. This indicates that the OAT math achievement is non-zero. The change in OATM is slightly proportional to MCA raw score, although not statistically significant as such, meaning that no immediate estimation of OAT-M standard score can be surmised based on a change in MCA raw score. Despite the lack of significance in the slope component, variability in this level-1 conditional model exhibited an decrease from
566.02 to 552.56 through the introduction of the predictor, MCA. Calculation of pseudo-$R^2$ ($\bar{R}^2$) would yield the amount of variance in the response variable contributed by predictor through comparison of the conditional model’s between subject variance to the previous unconditional or empty model’s between subject variance. This would permit the quantification of how much variability can be attributed by the inclusion of the predictor variable, MCA$_i$. The computation is as follows:

$$\bar{R}^2 = \frac{\tau^2_{00 \text{ [unconditional model]}} - \tau^2_{00 \text{ [conditional model]}}}{\tau^2_{00 \text{ [unconditional model]}}} = \frac{566.03 - 552.56}{566.03} = 0.024$$

This computation indicates that 2.4% of the variance between OAT mathematics achievement was decreased by the inclusion of the predictor variable, MCA. A decrease in variability suggests that MCA does partition a minimal amount of the OATM variability in performance, although to a very modest degree. In addition, this reveals that a large amount, totalling 97.6% of the variance, remains yet to be explained by other explanatory variables beyond MCA performance itself. Examination of $\chi^2$ for random effect variance components continued to be statistically significant, $\chi^2 (90) = 418.93, p < .01$. This finding offers evidence that there is considerable between-person variability in OAT math standard score results and that MCA failed to partition. With the introduction of additional predictor variables in the present conditional model, it is possible additional variability could be explained because MCA alone, was clearly insufficient, to account for sources of variation.
Conditional model for OAT reading. Production of the conditional model for the OAT reading subtest can be deconstructed in the exact manner as the above discussion. Like OAT mathematics, two fixed effects were estimated in this model for OAT reading, $\pi_{00}$ and $\pi_{10}$, that refer to the average reading achievement and change in OATR as partitioned by DIBELS ORF raw score. Through the introduction of the predictor variable, DIBELS, the average predicted reading achievement is 423.93 ($SE = 2.30$) and the change is is 0.20 ($SE = 0.05$) standard score points on the OAT reading subtest for every change in DIBELS raw scores. Unlike mathematics, both fixed effects were statistically significant, INTRCPT1, $\pi_{00}$, $t (70) = 184.65, p < .01$, and DIBELS, $\pi_{10}$, $t (282) = 4.22, p < .01$. These findings reveal that both reading achievement and the change in DIBELS raw scores were non zero and also that DIBELS both significantly predicts OATR and the relational strength. Level-1 variability reduced considerably with the introduction of the predictor variable from 519.25 to 334.82. Computation of pseudo-$R^2$ can be determined as follows:

$$\bar{R}^2 = \frac{(\tau^2_{00} \text{ [unconditional model]} - \tau^2_{00} \text{ [conditional model]})}{\tau^2_{00} \text{ [unconditional model]}}$$

$$= \frac{(519.25 - 334.82)}{519.25} = 0.355$$

This $\bar{R}^2$ value indicates that the addition of DIBELS accounted 35.5% of additional variability in OATR performance and potentially is a useful predictor of OATR performance overall. Examination of $\chi^2$ for random effect variance components continued to be statistically significant, $\chi^2 (70) = 571.45, p < .01$, positing that there continued to be
considerable variability in reading achievement and the change in OATR achievement standard scores. This furthers the expectation that including additional predictor variables could potentially explain more of this variability.

Towards Development of the Full Conditional Hierarchical Linear Models in OAT Reading and Math

Through this point, OAT reading and mathematics achievement subtest scores were first estimated through an empty model or one-way ANOVA. Next, these response variables were estimated with the introduction of a predictor variable of DIBELS ORF or MCA to consider how variance was partitioned, yielding evidence that both explanatory variables reduced variance in OAT standard score prediction. The final step in the successive model building process will be to explore how individual student level variables of race, gender, total school attendance, and participation in special education programming prospectively partitions out further additional elements of the variability in OAT subtest performance.

OAT mathematics full conditional model. Using the previous conditional model, the following individual student-level variables will be added as level-2 covariates: race [0 = White, 1 = Non-White], gender [0 = female, 1 = male], attendance [total days of attendance between 2002 and 2006], and special education affiliation [0 = none, 1 = special education student]. Results are summarized in Table 15:
Table 15

*Results of Full Conditional Hierarchical Linear Model: OAT Mathematics Subtest*

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Coefficient</th>
<th>Robust SE</th>
<th>t-ratio</th>
<th>df</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Average OATM Achievement Scores, $\pi_0i$ (intercept predictors)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTRCPT1, $\gamma_{00}$</td>
<td>402.47</td>
<td>2.32</td>
<td>173.21</td>
<td>86</td>
<td>0.000</td>
</tr>
<tr>
<td>GENDER, $\gamma_{01}$</td>
<td>4.30</td>
<td>4.54</td>
<td>0.95</td>
<td>86</td>
<td>0.347</td>
</tr>
<tr>
<td>RACE, $\gamma_{02}$</td>
<td>-16.05</td>
<td>5.68</td>
<td>-2.83</td>
<td>86</td>
<td>0.006</td>
</tr>
<tr>
<td>SPED, $\gamma_{03}$</td>
<td>-47.59</td>
<td>4.57</td>
<td>-10.41</td>
<td>86</td>
<td>0.000</td>
</tr>
<tr>
<td>TOT_ATEN, $\gamma_{04}$</td>
<td>0.20</td>
<td>0.04</td>
<td>4.91</td>
<td>86</td>
<td>0.000</td>
</tr>
<tr>
<td><strong>Change in OATM score, $\pi_1i$ (slope predictors)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTRCPT2, $\gamma_{10}$</td>
<td>-0.10</td>
<td>0.21</td>
<td>-0.46</td>
<td>172</td>
<td>0.648</td>
</tr>
<tr>
<td>GENDER, $\gamma_{11}$</td>
<td>0.48</td>
<td>0.46</td>
<td>1.08</td>
<td>172</td>
<td>0.283</td>
</tr>
<tr>
<td>RACE, $\gamma_{12}$</td>
<td>0.03</td>
<td>0.55</td>
<td>0.06</td>
<td>172</td>
<td>0.956</td>
</tr>
<tr>
<td>SPED, $\gamma_{13}$</td>
<td>0.27</td>
<td>0.66</td>
<td>0.43</td>
<td>172</td>
<td>0.667</td>
</tr>
<tr>
<td>TOT_ATEN, $\gamma_{14}$</td>
<td>0.00</td>
<td>0.00</td>
<td>1.11</td>
<td>172</td>
<td>0.268</td>
</tr>
</tbody>
</table>

*Note:* Restricted maximum likelihood estimation was used, sample size reduced to $n = 86$ due to run-time deletion by HLM 6.08 software. Deviance $= 1639.91$ with 2 parameters.

*Mean achievement in OAT mathematics.* Similar to the progressive development of the unconditional and conditional models above, a process of new
explanatory variables were fitted to the unconditional linear model as level-2 intercept-specific explanatory variables. These additions sought to describe how individual student race, gender, attendance, and special education participation could contribute to the partitioning of variance or prediction of OAT math subtest achievement. Several of these predictor variables yielded statistically significant results. First, racial differences were noted, $t(86) = -2.83, p < .01$. Second, participation in special education and related services also was significant for OATM prediction, $t(86) = -10.41, p < .01$. Finally, total attendance during the study timeframe was significantly related to OATM prediction, $t(86) = 4.91, p < .01$. Despite these significant findings, there were no gender differences in OATM performance, $t(86) = 0.95, p > .05$. Summarizing these data as a group, a mean OATM standard score of 402.47 ($SE = 2.32$) was obtained by participants within this study. Non-white students generally score 16.05 points below the mean OATM score ($SE = 5.16$), special education involved students score 47.59 points below the mean, and each day of absence from the average attendance rate can drop OATM scores by 0.20 ($SE = 0.04$) standard score points.

**Relational strength of moderator variables in OAT mathematics prediction.**

Across the four predictor variables of race, gender, special education, and attendance, none revealed statistically significant results towards influencing the strength or degree of relationship between those variables and OATM standard score prediction. That is, while the above results suggested that the moderator variables could change the predicted
standard score by a certain amount in 3 of the 4 moderator variables, none of the slope-specific moderator effects were significant to alter the OATM standard score.

**Pseudo $\bar{R}^2$: Proportion of variance explained in OAT math full model.** Using the conditional model build as a reference point, where MCA decreased variability, an estimation of the variance across both the intercept and slope-specific predictor variables can be calculated to determine how these additional covariates affected between subject variability:

$$\bar{R}^2 = \left( \tau_{00}^2 \text{ [conditional model]} - \tau_{00}^2 \text{ [full model]} \right) / \tau_{00}^2 \text{ [conditional model]}$$

$$= \left( 552.56 - 329.59 \right) / 552.56 = 0.404$$

The addition of the four student predictor variables partitioned out a sizable amount of variability of 40.4% beyond the 2.4% MCA alone accounted. Checking the model data to ensure no further level-2 predictor variables should be considered yields the following:

$$\text{ICC}_{level-2} = \sigma^2 / (\sigma^2 + \hat{\tau}^2)$$

$$= 292.47 / (292.47 + 301.46) = 0.49$$

This computation far exceeds the recommended threshold of $\geq .10$, revealing that subsequent predictor variables or covariates may need to be considered in the present model building process as there is sufficient variance remaining to partition in the presence of factors outside the present hierarchical linear model.
Covariance estimation for OAT math. The calculation of a reduction of error variance correlation, or covariance estimate, can be used to explain how those students who are higher levels of achievers modify their change in achievement over time, as follows:

$\hat{R}^2_B = \hat{\tau}^2_0 - \hat{\tau}^2_1 / \hat{\tau}^2_0$ where

$= ( 552.56 - 329.59 ) / 552.56 = 0.403$

This moderate correlation suggests that after controlling for the effects of race, gender, attendance, and special education participation, those students who exhibit the highest performance on the OAT math subtest are also those who display an increase in MCA performance between the fourth and fifth grades.

OAT reading full model. In the exact manner that the OAT mathematics full model was developed through the addition of the 4 explanatory variables (race, gender, attendance, and special education participation) with the conditional model on level-2, the current Fully Conditional OAT reading model was established. Results of the full OAT Reading model are presented in Table 16:
Table 16

Results of Conditional Full Hierarchical Linear Model: OAT Reading Subtest

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Coefficient</th>
<th>Robust SE</th>
<th>t-ratio</th>
<th>df</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average OATR Achievement Scores, $\pi_{0i}$ (intercept predictors)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTRCPT2, $\gamma_{00}$</td>
<td>423.77</td>
<td>2.12</td>
<td>199.42</td>
<td>66</td>
<td>0.000</td>
</tr>
<tr>
<td>GENDER, $\gamma_{01}$</td>
<td>-1.03</td>
<td>4.53</td>
<td>-0.23</td>
<td>66</td>
<td>0.820</td>
</tr>
<tr>
<td>RACE, $\gamma_{02}$</td>
<td>-7.45</td>
<td>4.68</td>
<td>-1.59</td>
<td>66</td>
<td>0.115</td>
</tr>
<tr>
<td>SPED, $\gamma_{03}$</td>
<td>-25.91</td>
<td>6.68</td>
<td>-3.89</td>
<td>66</td>
<td>0.000</td>
</tr>
<tr>
<td>TOT_ATEN, $\gamma_{04}$</td>
<td>0.11</td>
<td>0.08</td>
<td>1.30</td>
<td>66</td>
<td>0.200</td>
</tr>
<tr>
<td>Change in OATR Scores, $\pi_{1i}$ (slope predictors)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTRCPT2, $\gamma_{10}$</td>
<td>0.18</td>
<td>0.05</td>
<td>3.98</td>
<td>274</td>
<td>0.000</td>
</tr>
<tr>
<td>GENDER, $\gamma_{11}$</td>
<td>0.11</td>
<td>0.11</td>
<td>-1.02</td>
<td>274</td>
<td>0.311</td>
</tr>
<tr>
<td>RACE, $\gamma_{12}$</td>
<td>-0.02</td>
<td>0.11</td>
<td>-0.22</td>
<td>274</td>
<td>0.830</td>
</tr>
<tr>
<td>SPED, $\gamma_{13}$</td>
<td>-0.11</td>
<td>0.93</td>
<td>-1.20</td>
<td>274</td>
<td>0.233</td>
</tr>
<tr>
<td>TOT_ATEN, $\gamma_{14}$</td>
<td>0.00</td>
<td>0.00</td>
<td>0.79</td>
<td>274</td>
<td>0.429</td>
</tr>
</tbody>
</table>

Note: Restricted maximum likelihood estimation was used with runtime deletion to yield $n = 66$; deviance = 1624.85 with 2 parameters.

Mean achievement in OAT reading. A series of explanatory variables (race, gender, attendance, and special education participation) were added to the level-2 conditional equation. A significant relationship for those students who are participants in special education was found, $t (66) = -3.89, p < .01$. However, no other explanatory variables were statistically significant at $p < .05$, revealing that special education was the
most prominent influential factor for OATR subtest performance. Summarizing these data, the average OAT reading score of 423.77 ($SE = 2.12$) was obtained by participants in the present study. In contrast, students who receive special education services score, on average, 25.91 ($SE = 6.68$) standard score points lower than their peers on the OAT reading subtest.

**Relational strength of moderator variables in OAT reading prediction.** Unlike the influences each predictor variable exhibited on the average OATR standard score, there were no statistically significant relationships in predicting the relational strength or change in OATR scores. As a result, while the actual OATR standard score can be influenced by special education, for example, there is no evidence that the moderator variables of gender, race, special education or attendance do much to influence the nature of the relationship as quantified by the slope coefficients. However, for every increase in DIBELS raw score, there is a positive relationship of 0.18 ($SE = 0.04$) on OATR standard score.

**Pseudo $R^2$: Proportion of variance explained in OAT reading full model.** In establishing the conditional model as a baseline, the amount of variability reduced by the addition of the 4 explanatory factors can be estimated across through the following computation:

\[
\bar{R}^2 = \frac{(\tau^2_{00} \text{ [conditional model]} - \tau^2_{00} \text{ [full model]})}{\tau^2_{00} \text{ [conditional model]}} \\
= \frac{(334.82 - 293.12)}{334.82} = 0.125
\]
The addition of the 4 predictor variables accounted for an additional 12.5% of the variability in OATR standard score performance. When combined with DIBELS, these additional four predictor variables now account for 48% of the variability in OATR performance. Checking these data to ensure no further level-2 predictor variables should be considered yields the following:

\[
\text{ICC}_{\text{level-2}} = \frac{\sigma^2}{\sigma^2 + \hat{\tau}^2} \text{ where}
\]

\[
= \frac{187.73}{187.73 + 293.12} = 0.390
\]

This computation far exceeds the recommended threshold of \( \geq .10 \), revealing that subsequent predictor variables or covariates may need to be considered as there is sufficient variance remaining to partition in the presence of factors outside the present hierarchical linear model.

**Covariance estimation for OAT reading.** Similar to OAT mathematics, calculation of the covariance estimate can be used to explain how those students who are higher levels of achievers vary in the ability of DIBELS and the student predictor variables to predict OATR performance, as follows:

\[
\hat{R}^2_B = \hat{\tau}_0^2 - \hat{\tau}_1^2 = \frac{\hat{\tau}_0^2}{\hat{\tau}_0^2} \text{ where}
\]

\[
= \frac{(334.82 - 293.12)}{334.82} = 0.124
\]
The computation fails to establish a strong correlational relationship between OATR standard scores and the change in DIBELS when the variables of race, gender, special education and attendance are controlled.

**Hypothesis Testing: Using HLM for Inferential Statistical Analysis**

While statistical results and discussion of the hierarchical linear model construction were presented, the remainder of this present chapter will consider various inferential predictions concerning math and reading achievement consistent with the present study’s research questions.

**Research Questions and Hypotheses**

The below section serves to summarize the various research questions and hypotheses for the present study. The reader is reminded to review the preceding results chapter for further elaboration on the meaning of several of these findings.

**Level-one.** Question L1-1: Is there a significant predictive relationship between DIBELS oral reading fluency [ORF] and Ohio Achievement Test [OAT] reading subtest performance for elementary students?

Research hypothesis L1-1: There is a significant positive predictive relationship between a previous year’s DIBELS ORF measurement and a subsequent year’s OAT reading performance, (p<.05).
Null hypothesis L1-1: There is no significant predictive relationship between previous DIBELS ORF measurement and future OAT reading performance, at least at a statistically significant level, \((p \geq .05)\).

Result: Through consideration of the full hierarchical model for reading, with DIBELS as the level-1 explanatory variable, the average OATR reading was predicted to be 423.77 \((SE = 2.12)\). This finding was statistically significant, \(t (66) = 199.42, p < .01\), as was the slope-as-outcome analysis, \(t (274)=3.98, p < .01\). In addition, between person variability \((\tau^2)\) reduced from 519.25 to 334.82 while deviance, with 2 estimated parameters, reduced from 2450.79 to 2438.47. Consequently, the research hypothesis can be accepted and the null hypothesis rejected.

Question L1-2: Is there a significant predictive relationship between informal math concepts and applications [MCA] and OAT math subtest performance for elementary students?

Research hypothesis L1-2: There is a significant positive predictive relationship between a previous year’s MCA measurement and a subsequent year’s OAT math performance, \((p <.05)\).

Null hypothesis L1-2: There is no significant predictive relationship between previous MCA measurement and future OAT mathematical performance, at least a statistically significant level, \((p \geq .05)\).
Result: Similar to reading, by considering the full hierarchical linear model for math, with MCA as the level-1 explanatory variable, the average OATM standard score was predicted to be 402.47 (SE = 2.32). This finding was statistically significant, $t(86) = 173.21, p < .01$, although the slope-as-outcome analysis, $t(172) = -0.46, p > .05$, was not. Between person variability ($\tau^2$) reduced from 566.03 to 552.56 while deviance, with 2 estimated parameters, slightly increased from 1687.93 to 1690.93, revealing that MCA did slightly increase variability in the OATM prediction. With the non-significance of the slope t-test and increase in between person variability, the null hypothesis cannot be rejected.

**Level-two intercept related questions.** Answering the below intercept related research questions requires the use of the Fully Conditional Model for OATR and OATM, respectfully, and the addition of level-2 predictor variables of race, gender, attendance, and special education participation with the fixed effects of intercept. The intercept equations allow for the prediction of OAT reading and mathematics standard scores.

Question L2i-1: Does higher school attendance predict OAT reading/math subtest scores for students performing at the mean on DIBELS/MCA, *after controlling* for race, gender, and special education participants?
Research Hypothesis L2i-1: There is a significant predictive relationship between higher attending students and OAT reading/math subtest scores, after controlling for gender, race, and special education participation, \((p<.05)\).

Null hypothesis L2i-1: There is no significant predictive relationship between high attending students and OAT reading/math subtests, after controlling for gender, race, and special education participation, \((p\geq .05)\).

Results: In OAT reading achievement through the use of the fully conditional model, there was no significant effect for reading in the presence of the partial predictor variable, total attendance, \(t (66) = 1.29, p = 0.200\). Oppositely, in mathematics by using the fully conditional model, there is a statistically significant effect of total days of attendance departing from the mean attendance, \(t (86) =0.04, p <.01\). As a result, the research hypothesis for math and attendance can be accepted; revealing that as attendance increases from the average rate, there is a positive influence on OATM of .20 \((SE = 0.04)\) standard score points per day of attendance that departs from the grand mean. In contrast, one must fail to reject the null hypothesis for reading and attendance as there is no significant influence of total attendance on OAT reading standard scores.

Question L2i-2: Is there a significant predictive relationship of student gender to OAT reading/math subtest performance, for students performing at the mean on DIBELS/MCA, \textit{after controlling} for race, attendance, and special education participation?
Research hypothesis L2i-2: There is a significant positive predictive relationship of gender on OAT reading/math subtest performance after controlling for race, attendance, and special education participation, \((p<.05)\).

Null hypothesis L2i-2: There is no significant predictive relationship of gender on OAT reading/math subtest performance, after controlling for race, attendance, and special education participation, \((p\geq .05)\).

Results: Across both reading and mathematics, there are no statistically significant gender differences determined. Specifically, for math, \(t\) (86) = 0.95, \(p = 0.347\), while reading is \(t\) (66) = -0.228, \(p = 0.820\). Consequently, one must fail to reject both null hypotheses; that is to indicate there is no difference in OAT standard scores by gender across either reading or math.

Question L2i-3: Is there a significant predictive relationship between race and OAT reading/math subtest performance, for students performing at the mean on DIBELS/MCA, \textit{after controlling} for gender, attendance, and special education participation?

Research hypothesis L2i-3: There is a significant inverse predictive relationship between race and OAT reading/math subtest performance, after controlling for gender, attendance, and special education participation, \((p <.05)\).
Null hypothesis L2i-3: There is no significant predictive relationship between race and OAT reading/math subtest performance, after controlling for gender, attendance, and special education, (p≥ .05).

Results: In mathematics, there was a significant effect for race, \( t(86) = -2.83, p = 0.006 \). As a result, those students identified as “White” on their building enrollment forms scored, on average, 16.04 (SE = 5.68) standard score points higher than those classified as a part of any other racial group on the OAT mathematics subtest. In contrast, OAT reading was not as sensitive to racial differences as no significant differences were determined, \( t(66) = -1.59, p > .05 \). Consequently, the null hypothesis fails to be rejected for racial differences in reading but the research hypothesis accepted for racial differences in mathematics performance.

Question L2i-4: Is there a predictive relationship between participation in special education and OAT reading/math subtest performance, for students performing at the mean on DIBELS/MCA, after controlling for gender, race, and attendance?

Research hypothesis L2i-4: There is a significant inverse predictive relationship amongst special education participation and OAT reading/math subtest performance, after controlling for race, gender and attendance, (p <.05).
Null hypothesis L2i-4: There is no significant predictive relationship amongst special education participation and OAT reading/math subtest performance, after controlling for race, gender, and attendance, ($p \geq .05$)

Results: Participation in special education and related services was a significant predictor of both OAT reading and mathematics performance. In reading, $t (66) = -3.88$, $p = 0.00$ reveals the statistical significance while in math, $t (86) = -10.41$, $p = 0.00$ shows the same. When considered through the context of the predicted regression coefficients, special education recipients in OAT reading scored, on average, 25.91 ($SE = 6.67$) standard score points below those who do not receive special education. In mathematics, the impact was more austere, with those students identified as receiving special education scoring, on average, 47.59 ($SE = 4.57$) standard score points below their non-disabled peers.

**Level-two slope related questions.** Results for the slope-related research questions require the use of the Fully Conditional Model with all four explanatory variables added with the slope components added as a fixed effect.

Question L2s-1: Does the moderator variable of attendance rates for individual students influence the strength of DIBELS/MCA relationship with OAT reading/math subtest performance after controlling for gender, race, and special education participation?
Research hypothesis L2s-1: There is a strong moderator effect between the relationship of DIBELS/MCA and those students who have increased attendance, after controlling for race, gender, and special education participation, \((p < .05)\).

Null hypothesis L2s-1: There is no strong moderating relationship between DIBELS/MCA and those students who have increased attendance, after controlling for race, gender, and special education participation \((p \geq .05)\).

Results: Across both reading and mathematics, the influence of attendance rate does not significantly impact or change the nature of the relationship between the informal measure and the OAT; \(t_{\text{math}} (274) = 1.11, p > .05\) and \(t_{\text{reading}} (172) = -1.50, p > .05\). As a result, both null fail to be rejected, noting that attendance does not impact the nature of the relationship.

Question L2s-2: Does the moderator variable of student race influence the strength of DIBELS/MCA relationship to OAT reading/math subtest performance after controlling for attendance, gender, and special education participation?

Research hypothesis L2s-2: There is strong moderator effect between the relationship of DIBELS/MCA and race, after controlling for attendance, gender, and special education participation, \((p < .05)\).
Null hypothesis L2s-2: There is no strong moderating relationship between DIBELS/MCA and race, after controlling for attendance, gender, and special education participation, ($p \geq .05$).

Results: Racial differences are not present in the slope-component of either regression equation, $t_{\text{math}} (274) = -0.06, p > .05, t_{\text{reading}} (172) = -0.12, p > .05$. Given these findings, race does not appear to influence the nature of the relationship of predicting OAT performance in the presence of either informal measure. Consequently, the null hypotheses cannot fail to be rejected.

Question L2s-3: Does the moderator variable of gender influence the strength of relationship between DIBELS/MCA to OAT reading/math subtest performance after controlling for attendance, race, and special education participation?

Research hypothesis L2s-3: There is a strong moderator effect between the relationship of DIBELS/MCA and gender, after controlling for race, attendance, and special education participation, ($p < .05$).

Null hypothesis L2s-3: There is no significant moderating effect between DIBELS/MCA and gender, after controlling for race, attendance, and special education participation, ($p \geq .05$).
Results: Gender differences were not determined in either regression equation’s slope component, \( t_{\text{math}} (274) = 1.08, p > .05 \) and \( t_{\text{reading}} (172) = 0.76, p > .05 \). As a result, both null hypotheses failed to be rejected.

Question L2s-4: Does the moderator variable of participation in special education influence the strength of relationship on DIBELS/MCA with OAT reading/math subtest performance after controlling for attendance, race, and gender?

Research hypothesis L2s-4: There is a strong moderator effect between the relationship of DIBELS/MCA and special education, after controlling for race, gender, and attendance, (\( p < .05 \)).

Null hypothesis L2i-4: There is no significant moderating relationship between DIBELS/MCA and special education, after controlling for race, gender, and attendance, (\( p \geq .05 \)).

Results: Overall participation in special education fails to demonstrate an impact on the slope-components: \( t_{\text{math}} (274) = 0.43, p > .05 \) and \( t_{\text{reading}} (172) = 0.09, p > .05 \). Given such findings, both null hypotheses cannot be rejected.
CHAPTER 4: DISCUSSION

Introduction

Today’s educational climate is teeming with the establishment of national statewide academic content standards, alignment of numerous high stakes assessment instruments to such standards, development of national and local accountability measures, and assessment of school and teacher quality (Good, Simmons, & Kame'enui, 2001). From the 1983 Nation at Risk report (National Commission on Excellence in Education, 1983), there has been a resounding clamor for educational reform that has accelerated in intensity away from the ivory tower of higher education settings to mainstream citizenry. Inherent in these pushes on improving teacher and school performance, both through legislation and public attention, the utility of universal screening of school students has gained saliency. Schools are now charged with the responsibility to ensure that all students succeed, regardless of disability or existing skills, consistent with the most current reauthorization of ESEA, eponymously entitled No Child Left Behind. Efforts to identify struggling learners from their earliest arrival in their compulsory educational careers through a multi-tiered model of intervention endeavors that various multidimensional approaches of student assistance are available throughout the entire spectrum of academic and behavioral needs (Tilly, 2008).

The primacy of early reading and mathematics skill acquisition and development has been prominently acknowledged, particularly out of recognition that
each of those academic content areas holds such dominance in predicting future academic and emotional well-being for schoolchildren (National Assessment of Educational Progress, 2005). Innumerable literature reviews and meta-analytic studies have reflected on the relevance of reading and mathematics skills as foundations upon which subsequent academic success is predicated (e.g., Duncan et al., 2007).

Traditional methods to measure individual student success have only recently begun to take the initiative in developing appropriate research methodologies and statistical technologies to address the emergent emphasis on school, teacher, and student success. Inherent in the development of such approaches in the recurrent criticism in educational research that it has failed to attend the uniqueness of measuring students, nested within classrooms, which are subsequently and hierarchically organized within larger contextual units or repeated measurements that are nested within individual students that are subsequently nested in other macro-level contexts. Previous research typically would employ ordinary least squares regression and ANCOVA to predict these various nested structures, oftentimes with serious methodological flaws such as aggregation bias and misestimated precision (Kreft & de Leeuw, 1998). School psychologists, as one group of professionals who service schools and who are well trained in research, are particularly well suited to expand and implement these new models of analysis called hierarchical linear models that readily meet the data analysis complexities of both micro- and macro-level variables (Gelman & Hill, 2007).
The present study attempted to respond to these trends in school accountability, early identification of student academic skills, and evolving multilinear modeling methodologies. Through the use of a longitudinal data set, ranging in years between 2002 to 2006, elementary school students were evaluated with two categories of informal measurement, DIBELS oral reading fluency for early reading skill assessment (Fuchs & Fuchs, 2008), and math concepts and applications, a measure of composite mathematical skills beyond traditional computational accuracy (Fuchs, Hamlett, & Fuchs, 1998; Hirsch, 2006). Analysis through a methodology sensitive to contextual variation and repeated measurement of individual student units was employed to see how these informal measures may predict performance on a high-stakes measure of achievement, the Ohio Achievement Test in mathematics and reading, created by Ohio in response to the demands of NCLB (AIR Technical Team, 2008). Through the inclusion of various individual explanatory predictor variables, added to the level-2 model, including race, gender, attendance, and special education affiliation, the intent was to determine both how those informal measures may predict performance on a requisite achievement measure as well as perhaps influence change in academic achievement over time.

This present chapter will attempt to integrate this study's findings with previous literature, then conclude with a discussion of implications for practitioners, identified limitations within the current study, and conclude with suggestions for future research investigations.
General Implications of the Study and Alignment to Existing Literature

This current study attempted to employ hierarchical linear models to develop a two-fold model of prediction on the Ohio Achievement Test reading and mathematics subtests in the following two domains: (a) the amount of predictiveness partitioned by various individual student variables, including race, gender, attendance, and special education affiliation as assessed through repeated measurements of the DIBELS ORF and MCA nested within repeated measurements of individual students; and (b) estimation of magnitude of change on the Ohio Achievement Tests reading and mathematics subtests as contributed by individual student variables of race, gender, attendance, and special education participation.

The choice of the HLM methodology comes through contemporary literature citing the need for more advanced regression models and comparison techniques of residuals when such data are either nested within hierarchies (e.g., students within classrooms, within schools, and so forth) or when there are repeated measurements of some attribute over time that are nested within an individual student (Kreft & de Leeuw, 1998; Pedhazur, 1997; Raudenbush & Bryk, 2002). Previous attempts at such measurement suffer from methodological limitations that impinge upon the ability to consider these micro and macro-level contextual effects. The emergence of historic educational reform requiring individual states and school districts to establish measurements and comparison to national content standards further the urgency (Braden & Tayrose, 2008; Tilly, 2008). In addition, the arrival of multi-tiered models of
intervention now appreciate the diversity of student needs and recognize that early assessment and identification of prospectively struggling students increases the likelihood of ameliorating academic or behavioral deficits (Hirsch, 2006). No academic content areas are more salient than early literacy and mathematics skills as both have marked future predictive validity for academic resiliency and success (Fuchs & Fuchs, 2008).

**Reading.** Since 1989, Torgensen has forewarned that early reading failure predicts future reading and general academic poor performance throughout the remainder of a student’s educational career (e.g. Torgensen, 1998a; Torgensen, 1998b; Torgensen, Wagner, & Rashotte, 1997). In addition, the importance of intrinsic process oriented approaches of reading skills assessment over traditional IQ and predicted achievement discrepancies have also been asserted (e.g. Torgensen, 2001). Finally, consistent with current best practice approaches, continuous progress monitoring of student reading performance is recommended to identify those students at-risk or already struggling in reading (e.g. (Hosp & MacConnell, 2008).

Results of this study were consistent with these extant literature sources for the fidelity of informal reading assessment in predicting those students who will exhibit subsequent reading impairment in later years. By incorporating the informal measure, DIBELS oral reading fluency (Good & Kaminski, 2002), the amount of variability in a high stakes reading subtest, the OAT-R, was significantly partitioned out by 35.5 percent. In addition, identification as a special education student also yielded a significant hardship in future high stakes reading performance, further supporting Torgensen’s
claims of early reading difficulty predicating future reading difficulty. Concern over the efficacy of special education has long been discussed in the literature for various disability conditions (e.g. Fabiano et al., 2010), which this current study validates through acknowledgement of a 25.9 point disadvantage from the mean OAT-R performance. At the same time, consistent with technical documents and the psychometric properties of the OAT-R (AIR Technical Team, 2008), gender and racial differences were not overtly significantly observed. However, some pause should be noted in the impact of race as it was closer to statistical significance than other level-2 predictor variables, \( t \) (66) = -1.59, \( p = 0.12 \). This latter note may reflect that the small sample size reduced the power of the present study and a larger sample size could have, in fact, discerned racial differences.

Interestingly, the old adage of school attendance being a strong predictor of general achievement did not hold true for reading. Perhaps due to reading’s spiraling nature of skill development, those students who exhibited satisfactory reading skills in earlier years of schooling continued to maintain those skills sufficiently over time. Such results may lend further support to Torgensen’s perspective that poor readers stay poor readers while capable readers continue as such.

Slope as outcomes analysis, or moderator variables, in contrast, failed to yield any significance across any of the moderator variables. The continued significance of the DIBELS slope intercept, \( t \) (274) = 3.90, \( p < .01 \), suggests that departure from the grand mean of DIBELS raw score does impact the nature of OAT-R performance by a coefficient of 0.18.
Mathematics. The informal *Math Concepts and Applications* (Fuchs, Hamlett, & Fuchs, 1998) served little to impact the partitioning of variance in a high stakes mathematics achievement measure, the OAT-M subtest. Beginning with the unconditional or empty model, the inclusion of MCA only accounted for 2.4% of the variability, leaving another 97.6% of the variability accounted for through other factors. Such findings are contrary to current best practice recommendations of mathematical skills assessment (e.g. Fuchs & Fuchs, 2008) that emphasize a comprehensive assessment of numeracy skills. Likely the unique raw score structure of the MCA resulted in some contamination of these findings as the raw scores do not increase over time like DIBELS ORF and instead feature specific grade-level criterion cutoff points. Perhaps a measure that was exclusively a repeated measure of mathematical skills would yield a different result.

The inclusion of various level-2 predictor variables, including student demographic attributes such as race, gender, total attendance, and special education identification, served to yield a more significant provisioning of variance at 40.4%. Unlike reading, attendance was a highly significant predictor of mathematical performance, $t(86) = 4.91, p < .01$. Considering the necessity of continuous practice of mathematical skills and the spiraling nature of mathematical curricula (e.g. National Council of Teachers of Mathematics, 2000), this academic domain may be more sensitive to lack of instruction. Regardless of the reasoning, within the current study, for every day of student absence from the grand mean, 0.21 standard score points were lost on the
OAT-M. Racial differences were more pronounced within the present study, with non-White identified students scoring 16.04 points below those who were identified on school building enrollment forms as White American. These racial differences have been previously identified in the literature (e.g. Hall, Davis, Bolen, & Chia, 1999). Special education yielded evidence of the most significant disadvantage for OAT-M performance, with those students identified as receiving special education services scoring 47.59 standard score points below those who are not identified. Consistent with Hall et al. (1999)’s findings, gender differences were not discerned in this study, \( t (86) = 0.95, p > .05 \).

Unlike the slope as outcomes results using DIBELS, no significant moderator effects were noted across the four moderator variables or in the presence of MCA on the nature of the relationship with OAT-M. Subsequently, while the influence on the high stakes standard score values were readily apparent, the nature of the strength of the relationship remains to be determined.
Implications for Practitioners

In the context of this study’s findings, several relevant findings are immediately applicable to the daily school psychological practitioner. Perhaps most significant is the utility of informal assessment employing DIBELS Oral Reading Fluency as a quick progress monitoring measure of that vital academic skills consistent with at least one previous study (Vander Meer & Stollar, 2005). Within the current study, students were evaluated three times per annum, with only DIBELS yielding considerable predictive power in determining those students who are likely to succeed or fail on the looming specter of high-stakes achievement measures that impact district and school value-added statistics of educational effectiveness. Moreover, as many of these high-stakes tests start to carry onerous burdens to individual students such as advancement to subsequent grades or attainment of a high school diploma, identification of struggling students takes heightened priority before the students experience recurrent failure that could lead to negative self-appraisal of academic efficacy or even dropping out of school entirely. The usage of such informal measures provides a brief and efficient means towards identification of these students – a relevant consideration as teachers and educational personnel are continuously asked to do more with less resources, time, and ever-increasing curricular demands, inclusive of high-stakes test preparedness.

Despite mathematical assessment literature that promotes a comprehensive approach towards mathematical skill assessment (Fuchs, Hamlett, & Fuchs, 1998), the informal math concepts and applications failed to partition sufficient variability in
OAT-M performance. Potentially this is a consequence of the misalignment of MCA with Ohio’s math academic content standards as assessed on the OATM. Separately, because MCA does not represent a true repeated measure as the instrument differs by year, the psychometric properties may have introduced a notable confound in the present study. Clearly there are other more consequential predictors of OATM performance beyond those captured by this informal measure, including race, attendance and special education identification. Efforts to produce a measure more strongly aligned with the Ohio academic content standards may prove useful in yielding a measure of predictive validity as DIBELS ORF does in reading.

Identification as a student receiving special education services also represented a potent predictor across both OAT reading and mathematics subtests. Such a result makes rational sense given that a disability condition, by definition, is likely to adversely impact educational performance. However, despite the efforts of special education identification and services, these identified students do not appear to be as successful as their non-disabled peers in performing well on these measures. Whether such findings represent an issue of treatment integrity or fidelity, could not be determined with the context of this present investigation. Clearly, additional intervention efforts are needed to ensure that special education students are also equally successful on high-stakes testing, as already required under No Child Left Behind and IDEIA (Hirsch, 2006).

Finally, attendance appears to reveal a more potent influence on mathematics achievement than in reading. One may speculate that reading is an activity that occurs
throughout daily activities and can be partially supported through continuous efforts to improve reading skills. However, mathematics skills, requires the involvement of daily educational activities to ensure such skills are promoted due to the spiraling nature of mathematical skill acquisition (Fuchs & Fuchs, 2005). As a result, efforts to facilitate attendance may yield benefits towards ensuring students are successful in the high-stakes arena of mathematics achievement.

**Limitations of Present Study**

While several significant findings were determined, there still remains considerable variability to be partitioned in both reading and mathematics performance, particularly the latter, where MCA failed to account for only 2.4% of the relationship in math achievement. Such findings may reveal the presence of other unknown or undetermined predictor variables that were not considered in the present investigation. Even reading, while faring better with 35.5% of the variability acknowledged, still suggests that further explanatory variables are possible.

Separately, the current study was possibly compromised by a loss of the sample pool over the 2002 to 2006 longitudinal nature of this study due to attrition at the school building level. Such findings, while not an immediate consideration or research question, yield evidentiary documentation of the urban flight and increasing attrition experienced by urban school districts (Clotfelter, 2001). Individual variability in student’s attendance rates further diminished sample size, causing many students to miss individual administrations of the DIBELS and MCA assessments as well as the Ohio Achievement
Tests, causing unpaired and fragmented data. Because HLM cannot handle missing data at the second level of model building, any students missing one or more of their demographic data were eliminated. The irregularity of administration of the Ohio Achievement Test during its foundational years between 2003-2006 led to inconsistent data across the longitudinal sample where some students, due to absence or retention, were not given entirely equal opportunity to access these high-stakes achievement measures. Limitations of data analysis yielded another potent difficulty, as many data points were lost due to incomplete data, despite the ability of HLM to better accommodate missing level-1 data through estimation techniques (see Appendix C) than many other statistical methodologies. These factors conspired to require a modification in the original methodology that instead paired previous informal measurement to future high-stakes achievement results. The HLM 6.08 (Raudenbush, Byrk, & Congdon, 2009) analysis software also proved to be a daunting technology to use with concern raised that both the HLM software program and subsequent revisions resulted in differing analysis algorithms, introducing another potent limitation of general HLM analysis (Kreft & de Leeuw, 1998). Moreover, the technical requirements of HLM data structuring within the various statistics software led to wildly varying models of prediction, which could easily bias or obfuscate results without awareness on behalf of the researcher that such confounds were being introduced. Some assumptional violations were also made in the present HLM investigation. For example, in both reading and mathematics, level-1 heteroscedasticity of errors was found and violation of normality of data in OATR, third
grade, and attendance rates were noted that could compromise some aspects of predictiveness.

**Future Research Recommendations**

Future investigators are well advised to consider increasing the number of individual student cases to combat the issues of data attrition and missingness that inevitably result in the imperfect and non-experimentally rigid world of education where students are frequently absent or otherwise unavailable. Multicollinearity may be minimized by reducing the number of predictor variables and exploration of more sophisticated estimation techniques aside from the REML or ML employed by HLM 6.08 by default, discussed in Appendix C. Several interesting effects were also noted in the nested variable of school building, although they could not be readily explored in the current study, as many sources of variability by accounting for school could not be addressed such as unique classroom, teacher, and school climate effects. Subsequent HLM investigations may want to consider the design of this study in a 3-level model wherein repeated measurement of individual student performance indicators are nested within students, who are in turn, nested in larger contextual individual student variables and ultimately the school or district level. As an exploratory study, this investigation serves to ultimately establish a place for informal measurement of oral reading and global mathematical skills in their ability to predict subsequent performance on high-stakes achievement measures of the same academic content area. Noting that special education students are particularly at risk for inverse performance over time compared to their
non-disabled peers, further punctuates the saliency of early identification. Ultimately, resistance to the value-added momentum and quantification of school, teacher and student success does not have to represent a potent obstacle to educational reform; after all, teacher and school effectiveness remain as the most salient variables in determining whether children achieve success in school (Miller, 2008).
APPENDICES
APPENDIX A

Standardized Script for Administering DIBELS ORF and MCA
APPENDIX A:

Standardized Script for Administering DIBELS ORF and MCA

DIBELS Oral Reading Fluency Directions

The below instructions for DIBELS Oral Reading Fluency are presented to all volunteers approximately 30 minutes prior to the first administration during a pre-administration training session. Color-coded folders by grade are used for each set of probes and provided to each volunteer. A China marker, stopwatch and tissues are also provided to each volunteer. Volunteer examiners include school psychology trainees from local graduate training programs, district employed departmental school psychologists, and the building-level school psychologist who is responsible for overseeing the DIBELS administration process. Such examiners have each had extensive training in individualized assessment of children. Fidelity of administration is monitored during each administration by the presiding building-level school psychologist and when mathematically validated when entered into the Microsoft Excel spreadsheet.

General DIBELS ORF administration guidelines:

Each student will read three passages for one-minute each, color organized by grade level.

Errors are immediately verbally corrected by the examiner.

Assistance is given if a student has not offered a response for 3 seconds.

Repetitions are not scored as an error.

Insertions are not scored as an error.

Examiner’s verbatim instructions to student subject:

“I want to see how well you read. You will read three different passages for one-minute each. I want your best reading. If you make an error, I will correct you. If you do not know a word then I will tell it to you. Are you ready? Begin.”

Administration instructions:

At the end of 1-minute, please draw a vertical line on the student’s reading passage to indicate the last word read aloud. Calculate the total number of words read less the errors and record this number. You will record three numbers for each student. The median or middle score will be used for data entry into a Microsoft Excel spreadsheet that will
compute reliability intercorrelations and provide a printout for each classroom teacher usage.
Math Concepts and Applications Administration Instructions for Teachers

Unlike the DIBELS ORF, the MCA informal assessment is conducted in collaboration with the student’s individual classroom teachers. The below handout is disseminated to all teachers during a pre-administration staff meeting, coordinated by the building-level school psychologist. Results are also tabulated by the school psychologist who checks intercorrelations within a Microsoft Excel spreadsheet to confirm reliability of measurement.

Instructions to class room teachers:

The 6-minute math probe is a group-administered task. Please adhere to the following directions. This is the standardized method of administration, which has been used for 8 years. It is extremely important that you follow these directions so we may compare scores across school years and assist with our building predictions and concerns list.

Step 1. Pass out the papers and ask the students to neatly print their full name and room number. Once they have done this they are to put their pencils down and wait for the directions.

Step 2. Remind the students to keep their pencils on the desk. Then hold up a test and read the following as you show student that the test is comprised of 3 pages.

“This is a timed test. You will all have 6 minutes to do as many problems as you can. You will need to use your time wisely. Therefore, find the easy problems for you on page 1 (SHOW page 1), page 2 (SHOW page 2) and page 3 (SHOW page 3). Once you have done all the easy problems, go back to the beginning and try the more challenging ones. I will let you know when you have 3 minutes remaining. Ready? Begin.” (Immediately start timing and end at exactly 6 minutes).

Step 3. Please collect the papers in alphabetical order by last name and return to the school psychologist. Your building school psychologist will score these measures and provide you a spreadsheet of results.
APPENDIX B

Tabular Presentation of Level-1 and Level-2 Variables in Analysis
APPENDIX B:
Tabular Presentation of Level-1 and Level-2 Variables in Analysis

Table B1

*Level-I Variables of Analysis*

<table>
<thead>
<tr>
<th>Codes</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>STU_NBR</td>
<td>District unique 9 digit student identification number</td>
</tr>
<tr>
<td>OATR03</td>
<td>OAT Reading, 3&lt;sup&gt;rd&lt;/sup&gt; grade, spring administration only</td>
</tr>
<tr>
<td>OATR04</td>
<td>OAT Reading, 4&lt;sup&gt;th&lt;/sup&gt; grade</td>
</tr>
<tr>
<td>OATR05</td>
<td>OAT Reading, 5&lt;sup&gt;th&lt;/sup&gt; grade</td>
</tr>
<tr>
<td>OATR06</td>
<td>OAT Reading, 6&lt;sup&gt;th&lt;/sup&gt; grade</td>
</tr>
<tr>
<td>OATM05</td>
<td>OAT Math, 5&lt;sup&gt;th&lt;/sup&gt; grade</td>
</tr>
<tr>
<td>OATM06</td>
<td>OAT Math, 6&lt;sup&gt;th&lt;/sup&gt; grade</td>
</tr>
<tr>
<td>DIBELS&lt;sub&gt;xx&lt;/sub&gt;</td>
<td>DIBEL ORF spring administration raw score at xx grade</td>
</tr>
<tr>
<td>MCAS&lt;sub&gt;xx&lt;/sub&gt;</td>
<td>Math Concepts &amp; Application spring raw score at xx grade</td>
</tr>
<tr>
<td>INDEX&lt;sub&gt;x&lt;/sub&gt;</td>
<td>The index score number corresponding to each DIBELS/MCA administration</td>
</tr>
</tbody>
</table>
## Tabular Presentation of Level-2 Variables in Analysis

### Table B2

**Level-2 Variables of Analysis**

<table>
<thead>
<tr>
<th>Codes</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>STU_NBR</td>
<td>District unique 9 digit student identification number</td>
</tr>
<tr>
<td>GENDER</td>
<td>Biologic sex of individual, 0=female, 1=male</td>
</tr>
<tr>
<td>RACE</td>
<td>Racial identification of individual, 0=White, 1=Non-White</td>
</tr>
<tr>
<td>SCHOOL</td>
<td>School building of attendance, not used as nested variable</td>
</tr>
<tr>
<td>SPED</td>
<td>Student’s special education status, 0=No, 1=Eligible</td>
</tr>
<tr>
<td>TOT_ATEN</td>
<td>Total individual student attendance between 2002 and 2006</td>
</tr>
</tbody>
</table>

*Note.* Individual schools across building “A” and “B” were not nested as a part of the hierarchical model and instead used only for descriptive purposes. If included, the individual schools could have been represented in a level-3 equation in which repeated measures of DIBELS/MCA were nested within students, which would be nested within explanatory variables of race, gender, attendance, and SPED status, terminating in a top-level of analysis for each school building. This third level was not used because numerous contributions of variance at the school building level could have been present from classroom compositional dynamics, teacher style, to building culture as partial considerations. Such factors were not known or evaluated as a part of this present longitudinal investigation.
APPENDIX C

Missing, Inconsistent Data, and Study Variable Issues in HLM and The Present Study
APPENDIX C:

Missing, Inconsistent Data, and Study Variable Issues in HLM and The Present Study

Missing data is a tedious quandary that plagues most quantitative methodologies and statistical analyses (Gelman & Hill, 2007). Improper handling of missing datum values can lead to marked misrepresentation of the composite dataset and necessitates that the researcher assume that the absent data are analytically salient, unless proven otherwise. In other words, the reduction of sample size through such absence is of lesser importance than the prospect of bias or faulty generalization within the existing response set. In classical regression theory and contemporary statistical analysis software such as PASW, missing data are usually excluded through piecewise or listwise deletion in cases where any of the data points are missing, which represents a significant issue for single level models (Kreft & de Leeuw, 1998; Raudenbush & Bryk, 2002). Generally, removal of incomplete data sets, imputation, or modeling are methods to directly counteract the missing data condition using OLS, ANCOVA, MANOVA and various forms of regression in quantitative statistics.

What are missing and inconsistent data? Existing literature discussion and statistical examination of such data patterns. Problem solving the issue of missing data requires an exploration and consideration for the reasons the individual datum or pieces of data are missing in the first place. Measurement literature generally
acknowledges four types of “missingness mechanisms”: (a) missingness completely at random [MCAR]; (b) missingness at random [MAR]; (c) missingness that depends on unobserved predictors; and (d) missingness that depends on the missing value itself (Tabachnick & Fidell, 2007).

Data that is identified MCAR occurs if the probability of missingness is the same for all units; in other words, the missingness is randomly distributed across all possible observations. MCAR can be ascertained by dividing respondents into those with and without missing data, then employing \( t \)-tests of mean differences on key variables to establish whether the groups statistically significantly vary. Usually if data are MCAR, they may be discarded through listwise or pairwise deletion of cases without biasing the study’s inferential statistics. In the situation where data are not MCAR, imputation of data should be considered. Missingness at random [MAR] represents the probability that the missing data are not missing in a random manner but instead that the probability is increased that the missing data are randomly distributed within one or more subsamples. Typically, MAR is preferred for key explanatory variables with respect to the response variable and it is far more commonplace than MCAR. Assessment of MAR can be accomplished through separate variance \( t \)-tests in which rows that have 1% or more missing values, crossed with columns replete with all data, are compared. If the \( p < .05 \), the missing cases are significantly correlated with the column variable and prospectively not MAR. When an outcome variable is MAR, it is appropriate to exclude the missing cases to the extent that the regression controls for all the variables that affect the
probability of missingness. Otherwise, data imputation is suggested. Missingness that
depends on unobserved predictors occurs when information has not been obtained or
recorded with such missing information also predicting the missing data. An example of
such a condition might be a medical study involving a measurement of discomfort;
oftentimes, patients exhibiting discomfort because of a clinical trial treatment are more
likely to self-select themselves out of the study, thereby making their absence hardly
random and are significance to the study itself. Explicit modeling is necessary to
overcome potential bias of results in these circumstances. Finally, missingness that
depends on the missing value itself reflects a direct relationship between a missing data
point and the data itself. Such missingness mechanisms can be mitigated by increasing
the number of predictors in the missing-data model, which would elevate the data to
approximating missing data of the MAR subtype. Typically, these types of missing data
may be able to be predicted using other data in a researcher’s dataset but elevates the
prospective limitation that such predictive models may extrapolate beyond the range of
the originally observed data.

**Benefits of HLM for missing data.** HLM is considered to have a strategic
advantage to study growth patterns or repeated measures of a construct over time, as in
the current study, where the likelihood of considerable missing data has traditionally
plagued other longitudinal quantitative or level-one analytical approaches in the past.
Specifically, data that exhibit patterns of missingness of the MCAR and MAR subtypes
can be satisfactorily included in an HLM investigatory process through employing
maximum likelihood estimation (ML or MLE). Although a replete discussion of the strategies and mechanisms of data imputation far exceed this appendix entry, generically speaking, ML allows data to be interpreted as if no data are missing in the first place by incorporating data that have been observed at least once. In other words, having some component of the level-1 data available. Under the technical assumptions of ML, the data are treated as if they were MAR, assuming the researcher has validated through a priori analyses describe above, that there is no systematic reason for their missingness in the first place (Raudenbush & Bryk, 2002). Other methods of data imputation include pattern analysis, listwise and pairwise deletion, mean substitution, multiple regression by using non-missing data to predict values of missing data, approximate Bayesian bootstrap (ABB) methods, and multiple imputation. Generally, restricted maximum likelihood estimation (REML) and maximum likelihood (ML) estimation techniques produce similar regression coefficients but differ in terms of estimating variance components (Bryk & Raudenbush, 1992). If the number of level-2 units is small, ML variance estimates will be smaller than REML, possibility leading to arbitrarily short confidence intervals and biased significance tests. For additional information on these data estimation and imputation approaches, the reader is urged to consult Schafer (1999) for more elaborate treatment.

**Missing data and the current investigation.** Within the present study, considerable attrition of the longitudinal data set occurred as a function of figurative respondent mortality (Shaughnessy & Zechmeister, 1994) due to students withdrawing
from the buildings featured in the study, which would have compromised the ATTEND variable (level-2) and various level-1 variables such as OAT scores, DIBELS and MCA. Visual inspection of the data was suggestive of missingness depending on unobserved predictors and on missingness that depends on the missing value itself, as oftentimes there was an overt relationship apparent between absenteeism and missing OAT test scores. In addition, student achievement is related to retention as well as absenteeism, leading to students missing one or more administrations of specific OAT subtests also exhibiting corresponding lower academic skills. The tiered introduction of the OAT during 2003 onward further complicated OAT data collection as 34 students did not take the OAT math until 2006, despite a 2005 administration. However as a level-1 variable, OAT results could be handled through ML imputation approaches during the HLM analysis. Finally, the administration approach of the third grade OAT reading subtest itself represented another potential interrelationship and cause of data loss as students who achieved proficient status in the fall administration were excused from the spring re-administration. These factors converged to significantly abate the sample’s subject pool and required modification to the general methodology of the study as discussed in Chapter 2.

During the formal analysis stage, listwise deletion was employed through the HLM software at the second level of model computation, as HLM cannot handle missing data points during the level-2 aspect of analysis without producing an unrecoverable halting error. Remaining data had an intercorrelation matrix calculated and pairwise
deletion occurred as a process of initial model computation. A restricted maximum likelihood model of estimation (REML) was used as the last method to handle the missing data per the experimenter’s selection in the software options, which is also the default estimation method for HLM 6.08 (Raudenbush, Byrk, & Congdon, 2009). Consequently, the combination of participation attrition, co-relationships amongst missing data, and various elements of deletion resulted in a significant diminution of the sample that required formulation of a new methodology. Table 4 presents the erosion of the sample pool through general student withdrawal and itinerancy, which resulted in numerous deletions at level-2 of the HLM analysis due to missing SPED, ATTEND, GENDER, and RACE data. Remaining reductions of the sample through run-time deletion courtesy of the HLM software is notated within the various tables of statistics presented in the present study as appropriate for a given stage of model development.

Despite these considerations, the likelihood of sample erosion due to factors beyond the scope of the present study could exist, such as intrinsic student or familial factors that caused students to relocate multiple times, health, scholastic motivation, and endurance factors that caused students to be absent across all levels of assessment, and missing demographic values within the school district’s mainframe that yielded incomplete level-2 records. In these latter instances, such students were again removed due to constraints of the HLM software and its inability to handle missing level-2 data. Still, other explanations could abound and may represent a separate area of future investigation.
APPENDIX D

Raw HLM 6.08 Output for OAT-R:

Unconditional Model
APPENDIX D

Raw HLM 6.08 Output for OAT-R: Unconditional Model

Problem Title: no title

The data source for this run = reading.mdm
The command file for this run = C:\Users\ERICHM~1\AppData\Local\Temp\whltemp.hlm
Output file name = C:\Users\Erich Merkle\Documents\EM Dissertation Folder\HLM Files\Second Predefense HLM\hlm2.txt
The maximum number of level-1 units = 284
The maximum number of level-2 units = 71
The maximum number of iterations = 100
Method of estimation: restricted maximum likelihood

Weighting Specification

----------------------------------------
Weight Variable
Weighting? Name Normalized?
Level-1 no
Level-2 no
Precision no

The outcome variable is OATR

The model specified for the fixed effects was:

Level-1 Level-2
Coefficients Predictors

INTRCPT1, P0 INTRCPT2, B00
The model specified for the covariance components was:

---------------------------------------------------------

Sigma squared (constant across level-2 units)

Tau dimensions

INTRCPT1

Summary of the model specified (in equation format)

Level-1 Model

\[ Y = P0 + E \]

Level-2 Model

\[ P0 = B00 + R0 \]

Iterations stopped due to small change in likelihood function

******* ITERATION 2 *******

\[ \text{Sigma}_\text{_squared} = 177.73357 \]

\[ \text{Tau} \]

\[ \text{INTRCPT1,P0} = 519.25403 \]

\[ \text{Tau (as correlations)} \]

\[ \text{INTRCPT1,P0} = 1.000 \]

Random level-1 coefficient  Reliability estimate

| INTRCPT1, P0 | 0.921 |

The value of the likelihood function at iteration 2 = -1.225393E+003

The outcome variable is OATR
Final estimation of fixed effects:

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Coefficient</th>
<th>Error</th>
<th>T-ratio</th>
<th>d.f.</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRCPT1, P0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTRCPT2, B0</td>
<td>423.933099</td>
<td>2.817669</td>
<td>150.455</td>
<td>70</td>
<td>0.000</td>
</tr>
</tbody>
</table>

The outcome variable is OATR.

Final estimation of fixed effects (with robust standard errors):

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Coefficient</th>
<th>Error</th>
<th>T-ratio</th>
<th>d.f.</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRCPT1, P0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTRCPT2, B0</td>
<td>423.933099</td>
<td>2.797756</td>
<td>151.526</td>
<td>70</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Final estimation of variance components:

<table>
<thead>
<tr>
<th>Random Effect</th>
<th>Standard Deviation</th>
<th>Variance Component</th>
<th>df</th>
<th>Chi-square</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRCPT1, R0</td>
<td>22.78715</td>
<td>519.25403</td>
<td>70</td>
<td>888.02853</td>
<td>0.000</td>
</tr>
<tr>
<td>level-1, E</td>
<td>13.33168</td>
<td>177.73357</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Statistics for current covariance components model:

Deviance = 2450.786741
Number of estimated parameters = 2

Test of homogeneity of level-1 variance

Chi-square statistic = 131.92082
Number of degrees of freedom = 70
P-value = 0.000
APPENDIX E

Raw HLM 6.08 Output for OAT-R:

Conditional Model
APPENDIX E

Raw HLM 6.08 Output for OAT-R: Conditional Model

Program: HLM 6 Hierarchical Linear and Nonlinear Modeling
Authors: Stephen Raudenbush, Tony Bryk, & Richard Congdon
Publisher: Scientific Software International, Inc. (c) 2000
techsupport@ssicentral.com
www.ssicentral.com

-------------------------------------------------------------------------------
Module: HLM2R.EXE (6.08.29257.1)
-------------------------------------------------------------------------------

SPECIFICATIONS FOR THIS HLM2 RUN

Problem Title: no title

The data source for this run = reading.mdm
The command file for this run = C:\Users\ERICHM~1\AppData\Local\Temp\whlmtemp.hlm
Output file name = C:\Users\Erich Merkle\Documents\EM Dissertation Folder\HLM Files\Second Predefense HLM\hlm2.txt

The maximum number of level-1 units = 284
The maximum number of level-2 units = 71
The maximum number of iterations = 100
Method of estimation: restricted maximum likelihood

Weighting Specification

-----------------------------------
Weight
Variable

Weighting?   Name        Normalized?
Level-1      no
Level-2      no
Precision    no

The outcome variable is     OATR

The model specified for the fixed effects was:

-----------------------------------------------
Level-1                  Level-2
Coefficients             Predictors
-----------------------------------------------
INTRCPT1, P0             INTRCPT2, B00
#% DIBEL slope, P1       INTRCPT2, B10

'%' - The residual parameter variance for this level-1 coefficient has been set to zero.
'%' - This level-1 predictor has been centered around its grand mean.
The model specified for the covariance components was:
---------------------------------------------------------
Sigma squared (constant across level-2 units)

Tau dimensions
INTRCPT1

Summary of the model specified (in equation format)
---------------------------------------------------
Level-1 Model

\[ Y = P_0 + P_1 \times (DIBEL) + E \]

Level-2 Model

\[ P_0 = B_{00} + R_0 \]
\[ P_1 = B_{10} \]

Iterations stopped due to small change in likelihood function

******* ITERATION 8 ********

Sigma_squared = 185.99439

Tau

INTRCPT1, P0 334.82191

Tau (as correlations)

INTRCPT1, P0 1.000

----------------------------------------------------
Random level-1 coefficient  Reliability estimate
----------------------------------------------------
INTRCPT1, P0 0.878

The value of the likelihood function at iteration 8 = -1.219234E+003

The outcome variable is OATR
Final estimation of fixed effects:

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Standard Coefficient</th>
<th>Error</th>
<th>T-ratio</th>
<th>d.f.</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRCPT1, P0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTRCPT2, B00</td>
<td>423.933099</td>
<td>2.317732</td>
<td>182.909</td>
<td>70</td>
<td>0.000</td>
</tr>
<tr>
<td>DIBEL slope, P1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTRCPT2, B10</td>
<td>0.207800</td>
<td>0.041554</td>
<td>5.001</td>
<td>282</td>
<td>0.000</td>
</tr>
</tbody>
</table>

The outcome variable is OATR

Final estimation of fixed effects
(with robust standard errors)

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Standard Coefficient</th>
<th>Error</th>
<th>T-ratio</th>
<th>d.f.</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRCPT1, P0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTRCPT2, B00</td>
<td>423.933099</td>
<td>2.295893</td>
<td>184.648</td>
<td>70</td>
<td>0.000</td>
</tr>
<tr>
<td>DIBEL slope, P1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTRCPT2, B10</td>
<td>0.207800</td>
<td>0.049209</td>
<td>4.223</td>
<td>282</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Final estimation of variance components:

<table>
<thead>
<tr>
<th>Random Effect</th>
<th>Standard Deviation</th>
<th>Variance</th>
<th>df</th>
<th>Chi-square</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRCPT1, R0</td>
<td>18.29814</td>
<td>334.82191</td>
<td>70</td>
<td>571.45252</td>
<td>0.000</td>
</tr>
<tr>
<td>level-1, E</td>
<td>13.63798</td>
<td>185.99439</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Statistics for current covariance components model

Deviance = 2438.467486
Number of estimated parameters = 2

Test of homogeneity of level-1 variance

Chi-square statistic = 140.59276
Number of degrees of freedom = 70
P-value = 0.000
APPENDIX F

Raw HLM 6.08 Output for OAT-R:

Full Model
APPENDIX F

Raw HLM 6.08 Output for OAT-R: Full Model

Problem Title: no title

The data source for this run = reading.mdm
The command file for this run = C:\Users\ERICHM~1\AppData\Local\Temp\whlmtemp.hlm
Output file name = C:\Users\Erich Merkle\Documents\EM Dissertation Folder\HLM Files\Second Predefense HLM\hlm2.txt
The maximum number of level-1 units = 284
The maximum number of level-2 units = 71
The maximum number of iterations = 100
Method of estimation: restricted maximum likelihood

Weighting Specification

<table>
<thead>
<tr>
<th>Weighting?</th>
<th>Name</th>
<th>Normalized?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level-1</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>Level-2</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>Precision</td>
<td>no</td>
<td></td>
</tr>
</tbody>
</table>

The outcome variable is OATR
The model specified for the fixed effects was:

<table>
<thead>
<tr>
<th>Level-1 Coefficients</th>
<th>Level-2 Predictors</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRCPT1, P0</td>
<td>INTRCPT2, B00</td>
</tr>
<tr>
<td>$</td>
<td>GENDER, B01</td>
</tr>
<tr>
<td>$</td>
<td>RACE, B02</td>
</tr>
<tr>
<td>$</td>
<td>SPED, B03</td>
</tr>
<tr>
<td>$</td>
<td>TOT_ATEN, B04</td>
</tr>
<tr>
<td>#% DIBEL slope, P1</td>
<td>INTRCPT2, B10</td>
</tr>
<tr>
<td>$</td>
<td>GENDER, B11</td>
</tr>
<tr>
<td>$</td>
<td>RACE, B12</td>
</tr>
<tr>
<td>$</td>
<td>SPED, B13</td>
</tr>
<tr>
<td>$</td>
<td>TOT_ATEN, B14</td>
</tr>
</tbody>
</table>

'#$ - The residual parameter variance for this level-1 coefficient has been set to zero.

'%' - This level-1 predictor has been centered around its grand mean.

'S' - This level-2 predictor has been centered around its grand mean.

The model specified for the covariance components was:

Sigma squared (constant across level-2 units)

Tau dimensions

INTRCPT1

Summary of the model specified (in equation format)

Level-1 Model

\[ Y = P0 + P1 \times (DIBEL) + E \]

Level-2 Model

\[
\begin{align*}
P0 &= B00 + B01 \times (GENDER) + B02 \times (RACE) + B03 \times (SPED) + B04 \times (TOT_ATEN) + R0 \\
P1 &= B10 + B11 \times (GENDER) + B12 \times (RACE) + B13 \times (SPED) + B14 \times (TOT_ATEN)
\end{align*}
\]

Iterations stopped due to small change in likelihood function

***** ITERATION 10 *****
Sigma\_squared = 187.72870

Tau
INTRCPT1, P0  293.12151

Tau (as correlations)
INTRCPT1, P0  1.000

<table>
<thead>
<tr>
<th>Random level-1 coefficient</th>
<th>Reliability estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRCPT1, P0</td>
<td>0.862</td>
</tr>
</tbody>
</table>

The value of the likelihood function at iteration 10 = -1.215225E+003

The outcome variable is OATR

Final estimation of fixed effects:

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Coefficient</th>
<th>Error</th>
<th>T-ratio</th>
<th>d.f.</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRCPT2, B00</td>
<td>423.775257</td>
<td>2.253380</td>
<td>188.062</td>
<td>66</td>
<td>0.000</td>
</tr>
<tr>
<td>GENDER, B01</td>
<td>-1.033521</td>
<td>4.484287</td>
<td>-0.230</td>
<td>66</td>
<td>0.819</td>
</tr>
<tr>
<td>RACE, B02</td>
<td>-7.454739</td>
<td>4.549693</td>
<td>-1.639</td>
<td>66</td>
<td>0.106</td>
</tr>
<tr>
<td>SPED, B03</td>
<td>-25.912147</td>
<td>9.584244</td>
<td>-2.704</td>
<td>66</td>
<td>0.009</td>
</tr>
<tr>
<td>TOT_ATEN, B04</td>
<td>0.109083</td>
<td>0.095575</td>
<td>1.141</td>
<td>66</td>
<td>0.258</td>
</tr>
<tr>
<td>INTRCPT2, B10</td>
<td>0.182471</td>
<td>0.042659</td>
<td>4.277</td>
<td>274</td>
<td>0.000</td>
</tr>
<tr>
<td>GENDER, B11</td>
<td>-0.107012</td>
<td>0.087185</td>
<td>-1.227</td>
<td>274</td>
<td>0.221</td>
</tr>
<tr>
<td>RACE, B12</td>
<td>-0.023566</td>
<td>0.086797</td>
<td>-0.272</td>
<td>274</td>
<td>0.786</td>
</tr>
<tr>
<td>SPED, B13</td>
<td>-0.111514</td>
<td>0.163835</td>
<td>-0.681</td>
<td>274</td>
<td>0.496</td>
</tr>
<tr>
<td>TOT_ATEN, B14</td>
<td>0.001437</td>
<td>0.001829</td>
<td>0.786</td>
<td>274</td>
<td>0.433</td>
</tr>
</tbody>
</table>
The outcome variable is **OATR**

Final estimation of fixed effects
(with robust standard errors)

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>T-ratio</th>
<th>d.f.</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>For INTRCPT1, P0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTRCPT2, B0</td>
<td>423.775257</td>
<td>2.125073</td>
<td>199.417</td>
<td>66</td>
<td>0.000</td>
</tr>
<tr>
<td>GENDER, B01</td>
<td>-1.033521</td>
<td>4.532149</td>
<td>-0.228</td>
<td>66</td>
<td>0.820</td>
</tr>
<tr>
<td>RACE, B02</td>
<td>-7.454739</td>
<td>4.676851</td>
<td>-1.594</td>
<td>66</td>
<td>0.115</td>
</tr>
<tr>
<td>SPED, B03</td>
<td>-25.912147</td>
<td>6.672000</td>
<td>-3.884</td>
<td>66</td>
<td>0.000</td>
</tr>
<tr>
<td>TOT_ATEN, B04</td>
<td>0.109083</td>
<td>0.084244</td>
<td>1.295</td>
<td>66</td>
<td>0.200</td>
</tr>
<tr>
<td>For DIBEL slope, P1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTRCPT2, B10</td>
<td>0.182471</td>
<td>0.045842</td>
<td>3.980</td>
<td>274</td>
<td>0.000</td>
</tr>
<tr>
<td>GENDER, B11</td>
<td>-0.107012</td>
<td>0.105353</td>
<td>-1.016</td>
<td>274</td>
<td>0.311</td>
</tr>
<tr>
<td>RACE, B12</td>
<td>-0.023566</td>
<td>0.109596</td>
<td>-0.215</td>
<td>274</td>
<td>0.830</td>
</tr>
<tr>
<td>SPED, B13</td>
<td>-0.111514</td>
<td>0.093168</td>
<td>-1.197</td>
<td>274</td>
<td>0.233</td>
</tr>
<tr>
<td>TOT_ATEN, B14</td>
<td>0.001437</td>
<td>0.001813</td>
<td>0.792</td>
<td>274</td>
<td>0.429</td>
</tr>
</tbody>
</table>

Final estimation of variance components:

<table>
<thead>
<tr>
<th>Random Effect</th>
<th>Standard Deviation</th>
<th>Variance</th>
<th>df</th>
<th>Chi-square</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRCPT1, R0</td>
<td>17.12079</td>
<td>293.12151</td>
<td>66</td>
<td>467.13585</td>
<td>0.000</td>
</tr>
<tr>
<td>level-1, E</td>
<td>13.70141</td>
<td>187.72870</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Statistics for current covariance components model

- Deviance = 2430.450249
- Number of estimated parameters = 2

Test of homogeneity of level-1 variance

- Chi-square statistic = 152.81129
- Number of degrees of freedom = 70
- P-value = 0.000
APPENDIX G

Raw HLM 6.08 Output for OAT-M:

Unconditional Model
APPENDIX G

Raw HLM 6.08 Output for OAT-M: Unconditional Model

Program: HLM 6 Hierarchical Linear and Nonlinear Modeling
Authors: Stephen Raudenbush, Tony Bryk, & Richard Congdon
Publisher: Scientific Software International, Inc. (c) 2000
techsupport@ssicentral.com
www.ssicentral.com

-------------------------------------------------------------------------------
Module: HLM2R.EXE (6.08.29257.1)
-------------------------------------------------------------------------------

SPECIFICATIONS FOR THIS HLM2 RUN

Problem Title: no title

The data source for this run = math.mdm
The command file for this run = C:\Users\ERICHM~1\AppData\Local\Temp\whlmtemp.hlm
Output file name = C:\Users\Erich Merkle\Documents\EM Dissertation Folder\HLM Files\Second Predefense HLM\hlm2.txt

The maximum number of level-1 units = 182
The maximum number of level-2 units = 91
The maximum number of iterations = 100
Method of estimation: restricted maximum likelihood

Weighting Specification

--------------
Weight
Variable
Weighting?  Name   Normalized?
Level-1     no
Level-2     no
Precision   no

The outcome variable is OATM

The model specified for the fixed effects was:

---------------------------------------------
Level-1          Level-2
Coefficients    Predictors
---------------------------------------------
INTRCPT1, P0    INTRCPT2, B00
The model specified for the covariance components was:

---------------------------------------------------------
Sigma squared (constant across level-2 units)

Tau dimensions
  INTRCPT1

Summary of the model specified (in equation format)
---------------------------------------------------

Level-1 Model
  \[ Y = P0 + E \]

Level-2 Model
  \[ P0 = B00 + R0 \]

Iterations stopped due to small change in likelihood function

******* ITERATION 2 ******

Sigma_squared = 294.17033

Tau
  INTRCPT1,P0  566.02582

Tau (as correlations)
  INTRCPT1,P0  1.000

--------------------------------------------------
Random level-1 coefficient  Reliability estimate
--------------------------------------------------
  INTRCPT1, P0  0.794

--------------------------------------------------

The value of the likelihood function at iteration 2 = -8.439655E+002
The outcome variable is OATM

Final estimation of fixed effects:

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Coefficient</th>
<th>Error</th>
<th>T-ratio</th>
<th>d.f.</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRCPT1, P0</td>
<td>402.489011</td>
<td>2.799354</td>
<td>143.779</td>
<td>90</td>
<td>0.000</td>
</tr>
<tr>
<td>INTRCPT2, B00</td>
<td>402.489011</td>
<td>2.783931</td>
<td>144.576</td>
<td>90</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Final estimation of fixed effects (with robust standard errors)

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Coefficient</th>
<th>Error</th>
<th>T-ratio</th>
<th>d.f.</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRCPT1, P0</td>
<td>402.489011</td>
<td>2.783931</td>
<td>144.576</td>
<td>90</td>
<td>0.000</td>
</tr>
<tr>
<td>INTRCPT2, B00</td>
<td>402.489011</td>
<td>2.783931</td>
<td>144.576</td>
<td>90</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Final estimation of variance components:

<table>
<thead>
<tr>
<th>Random Effect</th>
<th>Standard Deviation</th>
<th>Variance Component</th>
<th>df</th>
<th>Chi-square</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRCPT1, R0</td>
<td>23.79130</td>
<td>566.02582</td>
<td>90</td>
<td>436.34577</td>
<td>0.000</td>
</tr>
<tr>
<td>level-1, E</td>
<td>17.15139</td>
<td>294.17033</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Statistics for current covariance components model

Deviance = 1687.930926

Number of estimated parameters = 2

Test of homogeneity of level-1 variance

Chi-square statistic = 137.23597

Number of degrees of freedom = 90

P-value = 0.001
APPENDIX H

Raw HLM 6.08 Output for OAT-M:

Conditional Model
APPENDIX H

Raw HLM 6.08 Output for OAT-M: Conditional Model

Program: HLM 6 Hierarchical Linear and Nonlinear Modeling
Authors: Stephen Raudenbush, Tony Bryk, & Richard Congdon
Publisher: Scientific Software International, Inc. (c) 2000
techsupport@ssicentral.com
www.ssicentral.com

-------------------------------------------------------------------------------
Module: HLM2R.EXE (6.08.29257.1)
-------------------------------------------------------------------------------

SPECIFICATIONS FOR THIS HLM2 RUN

Problem Title: no title
The data source for this run = math.mdm
The command file for this run = C:\Users\ERICHM~1\AppData\Local\Temp\whlmtemp.hlm
Output file name = C:\Users\Erich Merkle\Documents\EM Dissertation Folder\HLM Files\Second Predefense HLM\hlm2.txt
The maximum number of level-1 units = 182
The maximum number of level-2 units = 91
The maximum number of iterations = 100
Method of estimation: restricted maximum likelihood

Weighting Specification

<table>
<thead>
<tr>
<th>Weight Variable</th>
<th>Weighting?</th>
<th>Name</th>
<th>Normalized?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level-1</td>
<td>no</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level-2</td>
<td>no</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Precision</td>
<td>no</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The outcome variable is OATM

The model specified for the fixed effects was:

<table>
<thead>
<tr>
<th>Level-1 Coefficients</th>
<th>Level-2 Predictors</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRCPT1, P0</td>
<td>INTRCPT2, B00</td>
</tr>
<tr>
<td>%# MCA slope, P1</td>
<td>INTRCPT2, B10</td>
</tr>
</tbody>
</table>

'%' - This level-1 predictor has been centered around its grand mean.
The model specified for the covariance components was:

---------------------------------------------
Sigma squared (constant across level-2 units)

Tau dimensions
    INTRCPT1

Summary of the model specified (in equation format)
---------------------------------------------

Level-1 Model

\[ Y = P0 + P1 \times (MCA) + E \]

Level-2 Model

\[ P0 = B00 + R0 \]
\[ P1 = B10 \]

Iterations stopped due to small change in likelihood function

******* ITERATION 21 *******

Sigma_squared = 301.46491

Tau
    INTRCPT1,P0  552.55836

Tau (as correlations)
    INTRCPT1,P0  1.000

---------------------------------------------
Random level-1 coefficient  Reliability estimate
---------------------------------------------
    INTRCPT1, P0  0.786

---------------------------------------------

The value of the likelihood function at iteration 21 = -8.454628E+002
The outcome variable is OATM

Final estimation of fixed effects:

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Coefficient</th>
<th>Error</th>
<th>T-ratio</th>
<th>d.f.</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRCPT1, P0</td>
<td>402.489011</td>
<td>2.780052</td>
<td>144.777</td>
<td>90</td>
<td>0.000</td>
</tr>
<tr>
<td>INTRCPT2, B00</td>
<td>0.051959</td>
<td>0.221575</td>
<td>0.234</td>
<td>180</td>
<td>0.815</td>
</tr>
</tbody>
</table>

The outcome variable is OATM

Final estimation of fixed effects (with robust standard errors)

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Coefficient</th>
<th>Error</th>
<th>T-ratio</th>
<th>d.f.</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRCPT1, P0</td>
<td>402.489011</td>
<td>2.761423</td>
<td>145.754</td>
<td>90</td>
<td>0.000</td>
</tr>
<tr>
<td>INTRCPT2, B10</td>
<td>0.051959</td>
<td>0.224588</td>
<td>0.231</td>
<td>180</td>
<td>0.817</td>
</tr>
</tbody>
</table>

Final estimation of variance components:

<table>
<thead>
<tr>
<th>Random Effect</th>
<th>Standard Deviation</th>
<th>Variance</th>
<th>df</th>
<th>Chi-square</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRCPT1, R0</td>
<td>23.50656</td>
<td>552.55836</td>
<td>90</td>
<td>418.93033</td>
<td>0.000</td>
</tr>
<tr>
<td>level-1,</td>
<td>17.36274</td>
<td>301.46491</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Statistics for current covariance components model

Deviance = 1690.925662
Number of estimated parameters = 2

Test of homogeneity of level-1 variance

Chi-square statistic = 139.59426
Number of degrees of freedom = 90
P-value = 0.001
APPENDIX I

Raw HLM 6.08 Output for OAT-M:

Full Model
APPENDIX I

Raw HLM 6.08 Output for OAT-M: Full Model

Problem Title: no title

The data source for this run = math.mdm
The command file for this run = C:\Users\ERICHM~1\AppData\Local\Temp\whltemp.hlm
Output file name = C:\Users\Erich Merkle\Documents\EM Dissertation Folder\HLM Files\Second Predefense HLM\hlm2.txt
The maximum number of level-1 units = 182
The maximum number of level-2 units = 91
The maximum number of iterations = 100
Method of estimation: restricted maximum likelihood

Weighting Specification

<table>
<thead>
<tr>
<th>Weighting?</th>
<th>Variable</th>
<th>Normalized?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level-1</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>Level-2</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>Precision</td>
<td>no</td>
<td></td>
</tr>
</tbody>
</table>

The outcome variable is OATM
The model specified for the fixed effects was:

----------------------------------------------------
<table>
<thead>
<tr>
<th>Level-1</th>
<th>Level-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficients</td>
<td>Predictors</td>
</tr>
<tr>
<td>----------------</td>
<td>---------------</td>
</tr>
<tr>
<td>INTRCPT1, P0</td>
<td>INTRCPT2, B00</td>
</tr>
<tr>
<td>$</td>
<td>GENDER, B01</td>
</tr>
<tr>
<td>$</td>
<td>RACE, B02</td>
</tr>
<tr>
<td>$</td>
<td>SPED, B03</td>
</tr>
<tr>
<td>$</td>
<td>TOT_ATEN, B04</td>
</tr>
<tr>
<td>$</td>
<td>MCA slope, P1</td>
</tr>
<tr>
<td>$</td>
<td>INTRCPT2, B10</td>
</tr>
<tr>
<td>$</td>
<td>GENDER, B11</td>
</tr>
<tr>
<td>$</td>
<td>RACE, B12</td>
</tr>
<tr>
<td>$</td>
<td>SPED, B13</td>
</tr>
<tr>
<td>$</td>
<td>TOT_ATEN, B14</td>
</tr>
</tbody>
</table>

'#' - The residual parameter variance for this level-1 coefficient has been set to zero.

'%' - This level-1 predictor has been centered around its grand mean.

'S' - This level-2 predictor has been centered around its grand mean.

The model specified for the covariance components was:

---------------------------------------------------------
| Sigma squared (constant across level-2 units) |
| Tau dimensions                                   |
| INTRCPT1                                        |

Summary of the model specified (in equation format)

Level-1 Model

\[ Y = P0 + P1 \cdot (MCA) + E \]

Level-2 Model

\[ P0 = B00 + B01 \cdot (GENDER) + B02 \cdot (RACE) + B03 \cdot (SPED) + B04 \cdot (TOT_ATEN) + R0 \]
\[ P1 = B10 + B11 \cdot (GENDER) + B12 \cdot (RACE) + B13 \cdot (SPED) + B14 \cdot (TOT_ATEN) \]

Iterations stopped due to small change in likelihood function

******* ITERATION 16 *******
Sigma_squared = 292.46844

Tau
INTRCPT1,P0  329.58706

Tau (as correlations)
INTRCPT1,P0  1.000

----------------------------------------------------
Random level-1 coefficient   Reliability estimate
----------------------------------------------------
INTRCPT1, P0                        0.693
----------------------------------------------------

The value of the likelihood function at iteration 16 = -8.199557E+002

The outcome variable is OATM

Final estimation of fixed effects:

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Coefficient</th>
<th>Error</th>
<th>T-ratio</th>
<th>d.f.</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRCPT2, B00</td>
<td>402.470154</td>
<td>2.381540</td>
<td>168.996</td>
<td>86</td>
<td>0.000</td>
</tr>
<tr>
<td>GENDER, B01</td>
<td>4.301380</td>
<td>4.661130</td>
<td>0.923</td>
<td>86</td>
<td>0.359</td>
</tr>
<tr>
<td>RACE, B02</td>
<td>-16.047275</td>
<td>5.182136</td>
<td>-3.097</td>
<td>86</td>
<td>0.003</td>
</tr>
<tr>
<td>SPED, B03</td>
<td>-47.589616</td>
<td>9.222710</td>
<td>-5.160</td>
<td>86</td>
<td>0.000</td>
</tr>
<tr>
<td>TOT_ATEN, B04</td>
<td>0.205375</td>
<td>0.071658</td>
<td>2.866</td>
<td>86</td>
<td>0.006</td>
</tr>
</tbody>
</table>

For MCA slope, P1

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Coefficient</th>
<th>Error</th>
<th>T-ratio</th>
<th>d.f.</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRCPT2, B10</td>
<td>-0.105197</td>
<td>0.219890</td>
<td>-0.478</td>
<td>172</td>
<td>0.633</td>
</tr>
<tr>
<td>GENDER, B11</td>
<td>0.485964</td>
<td>0.441085</td>
<td>1.102</td>
<td>172</td>
<td>0.273</td>
</tr>
<tr>
<td>RACE, B12</td>
<td>0.031319</td>
<td>0.473996</td>
<td>0.066</td>
<td>172</td>
<td>0.948</td>
</tr>
<tr>
<td>SPED, B13</td>
<td>0.267739</td>
<td>0.722959</td>
<td>0.370</td>
<td>172</td>
<td>0.711</td>
</tr>
<tr>
<td>TOT_ATEN, B14</td>
<td>0.002369</td>
<td>0.004644</td>
<td>0.510</td>
<td>172</td>
<td>0.610</td>
</tr>
</tbody>
</table>
The outcome variable is OATM

Final estimation of fixed effects
(with robust standard errors)

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>T-ratio</th>
<th>d.f.</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>For INTRCPT1, P0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTRCPT2, B00</td>
<td>402.470154</td>
<td>2.323615</td>
<td>173.209</td>
<td>86</td>
<td>0.000</td>
</tr>
<tr>
<td>GENDER, B01</td>
<td>4.301380</td>
<td>4.544944</td>
<td>0.946</td>
<td>86</td>
<td>0.347</td>
</tr>
<tr>
<td>RACE, B02</td>
<td>-16.047275</td>
<td>5.678946</td>
<td>-2.826</td>
<td>86</td>
<td>0.006</td>
</tr>
<tr>
<td>SPED, B03</td>
<td>-47.589616</td>
<td>4.571889</td>
<td>-10.409</td>
<td>86</td>
<td>0.000</td>
</tr>
<tr>
<td>TOT_ATEN, B04</td>
<td>0.205375</td>
<td>0.041866</td>
<td>4.905</td>
<td>86</td>
<td>0.000</td>
</tr>
<tr>
<td>For MCA slope, P1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTRCPT2, B10</td>
<td>-0.105197</td>
<td>0.230298</td>
<td>-0.457</td>
<td>172</td>
<td>0.648</td>
</tr>
<tr>
<td>GENDER, B11</td>
<td>0.485964</td>
<td>0.451161</td>
<td>1.077</td>
<td>172</td>
<td>0.283</td>
</tr>
<tr>
<td>RACE, B12</td>
<td>0.031319</td>
<td>0.563404</td>
<td>0.056</td>
<td>172</td>
<td>0.956</td>
</tr>
<tr>
<td>SPED, B13</td>
<td>0.267739</td>
<td>0.622803</td>
<td>0.430</td>
<td>172</td>
<td>0.667</td>
</tr>
<tr>
<td>TOT_ATEN, B14</td>
<td>0.002369</td>
<td>0.002130</td>
<td>1.112</td>
<td>172</td>
<td>0.268</td>
</tr>
</tbody>
</table>

Final estimation of variance components:

<table>
<thead>
<tr>
<th>Random Effect</th>
<th>Standard Deviation</th>
<th>Variance</th>
<th>df</th>
<th>Chi-square</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRCPT1, R0</td>
<td>18.15453</td>
<td>329.58706</td>
<td>86</td>
<td>276.75285</td>
<td>0.000</td>
</tr>
<tr>
<td>level-1, E</td>
<td>17.10171</td>
<td>292.46844</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Statistics for current covariance components model

| Deviance                   | 1639.911400        |
| Number of estimated parameters | 2                  |

Test of homogeneity of level-1 variance

| Chi-square statistic       | 153.93429          |
| Number of degrees of freedom | 90                |
| P-value                    | 0.000              |


