2005

The Effects of Developmentally Appropriate Practices on Children's Reading Development from Kindergarten Through Third Grade

Alper T. Kumtepe

THE FLORIDA STATE UNIVERSITY

COLLEGE OF EDUCATION

THE EFFECTS OF DEVELOPMENTALLY APPROPRIATE PRACTICES ON CHILDREN'S READING DEVELOPMENT FROM KINDERGARTEN THROUGH THIRD GRADE

By

ALPER T. KUMTEPE

A Dissertation submitted to
Department of Childhood Education, Reading and Disability Services
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Degree Awarded:
Fall Semester, 2005

The members of the Committee approve the dissertation of Alper T. Kumtepe defended on October 28, 2005.

Ithel Jones
Professor Directing Dissertation

Akihito Kamata
Outside Committee Member

Charles H. Wolfgang
Committee Member

Vickie Lake
Committee Member

Approved:

Ithel Jones, Chair, Department of Childhood Education, Reading, and Disability Services

The Office of Graduate Studies has verified and approved the above named committee members.

To my family

ACKNOWLEDGEMENTS

There are many people who contributed and supported the writing of this dissertation. Grateful appreciation is extended to my major professor Dr. Ithel Jones, without whose guidance, patience, and encouragement, this study would never have been completed. He believed in my abilities and created the most positive environment to conduct this work. My sincere thanks also go to Dr. Charles H. Wolfgang and Dr. Vickie Lake for their most helpful criticism in this study as members of my committee. I am particularly grateful to Dr. Akihito Kamata for his invaluable expertise and methodological assistance.

My appreciation is extended to my friends Ummuhan Yesil Dagli, Burhanettin Keskin, and Dina Vyortkina for their willingness to help and genuine interest they showed during the study.

My greatest debt is to my family, my mother Ayla Kumtepe, my father Hidayet Kumtepe, and my little brother Onur Kumtepe who have believed in and supported me throughout this endeavor. During these years away from home, I always felt their unconditional love and continuous encouragement. I also would like to acknowledge the moral support of my other family members.

Most importantly, my appreciation goes to my love Evrim Genc, without whom I would not have completed this study. She always provided her friendship, support, tireless encouragement, humor, and understanding during the years we both suffered from being away from our families. Evrimcan, you made these years more fun and livable. It has always been good to come *home* because of you.

Finally, I express my sincere appreciation to the Turkish Ministry of Education for the scholarship that made my graduate studies possible. I will always try to live up to the expectations of this great opportunity they provided me with.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

This study was an attempt to investigate the effects of developmentally appropriate classroom practices (DAP) on children's reading achievement from kindergarten through third grade. Considering the limitations of previous research methodology, an appropriate multilevel analysis approach was employed. Furthermore, the differential effects of DAP as a function of children's age, gender, family SES, ethnicity, and school sector was examined.

The results of multilevel analyses illustrated a consistent pattern in assessing the effects of DAP. Whereas DAP did not have any significant effects on a standardized direct measure of reading, there were significant and positive effects detected at all three grades on an indirect non-standardized indirect measure of reading. The findings contribute to the existing literature in understanding the relationship between DAP and student outcomes by differentiating the effects of DAP on two measures that are different in nature. Therefore, decisions about the success of developmentally appropriate practices must consider multiple indicators of student achievement. Furthermore, evidence for differential effects of DAP was not enough to conclude that DAP was more beneficial for certain groups of children. In particular, there were no systematic patterns of racial differences in children's reading development as a function of DAP.

The implications for early childhood assessment and classroom practices are presented in addition to suggestions for future research efforts.

CHAPTER I
INTRODUCTION

Since the National Association for the Education of Young Children's (NAEYC) first published the position statement on developmentally appropriate practices (DAP) in 1987 (Bredekamp, 1987), revised in 1997 by Bredekamp and Copple (1997), the concept of developmental appropriateness has extensively influenced both research and practice in the field of early childhood education. It has been adopted and employed by many school districts and individual schools, and also numerous studies have examined the topic. Although the guidelines initiated a number of controversies, confusions, and myths, they have been accepted as the consensus starting point for the discussion of DAP (Jones & Gullo, 1999; Walsh, 1991).

Preparation of DAP was a counter effort to trends toward using formal academic curricula and didactic teaching practices in preschools and kindergartens. Based on Piaget's theory of the construction of knowledge as the basis for the cognitive-developmental approach (DeVries, 1997; Kamii & DeVries, 1993), DAP guidelines rely upon the principle that children construct their own knowledge through interactions with the social and physical environment. The definition of developmentally appropriate practice (Bredekamp, 1987; Bredekamp & Copple, 1997) argues that young children should be exposed to learning experiences that are in harmony with their current levels of development in order to aid and support their learning. DAP utilizes an approach to the education of young children that focuses on the child as a developing human being and life-long learner. The child is viewed as an active participant in the learning process; a participant who constructs meaning and knowledge through interaction with others, friends and family, materials and environment (Bredekamp, 1987; Bredekamp & Copple, 1997; Getswicki, 1999; Goffin & Wilson, 2001).

Since the publication of the DAP statement (Bredekamp, 1987), the construct of developmental appropriateness has been investigated by educational researchers. Several studies have examined the effects of DAP on cognitive outcomes (e.g., Burts et al., 1993; Dunn, Beach, & Kontos, 1994; Hirsh-Pasek, Hyson, & Rescorla, 1990; Huffman & Speer, 2000; Jones & Gullo, 1999), whereas others studied social and emotional outcomes (e.g., Burts, Hart, & Kirk, 1990; Jones & Gullo, 1999; Marcon, 1992; Stipek, Feiler, Daniels, & Milburn, 1995) affected by the construct.

Studies on language development in child-initiated and academically oriented programs reported that developmentally appropriate programs were associated with better language outcomes. Marcon (1992) stated that children in child-initiated classrooms had better verbal skills than children in academically focused programs. In another study, children's receptive language was better when children were exposed to developmentally appropriate activities in programs with a higher quality literacy environment (Dunn, Beach, & Kontos, 1994). Furthermore, children's perceptions about their own cognitive competence were more positive when they attended child-initiated programs rather than traditional academically oriented programs (Mantizocopoulos, Neuhart-Pritchett, & Morelock, 1994; Stipek, et al., 1995). However, another group of studies reported mixed or no effects (Hirsh-Pasek et al., 1990; Huffman & Speer, 2000; Jones & Gullo, 1999; Stipek et al., 1998; Van Horn & Ramey, 2003), and one study reported negative effects (Stipek et al., 1995).

Despite the growing number of empirical research on the topic, evidently, there are still potential questions to be explored regarding the effectiveness of DAP in the context of early childhood education. In particular, the need for investigating longitudinal effects of kindergarten DAP on children's academic achievement has been addressed by researchers in the literature (Burts et al., 1993; Jones & Gullo, 1999; Van Horn & Ramey, 2003).

This study was an attempt to explore the effects of developmentally appropriate classroom practices provided on children's academic achievement, namely reading scores, through the first four years of schooling. Considering the limitations of previous research methodology (Van Horn & Ramey, 2003), such as the violations of the assumptions of independence of observations that is essential to regression and ANOVA models (Tabachnick & Fidel, 1996), the proposed study employed an appropriate multilevel analyses that considers the nested data structure to investigate the extent to which DAP is associated with children's reading achievement from kindergarten through 3rd grade. Furthermore, differential effects of DAP as a function of children's age, genders, family SES, ethnicity, and school sector were examined.

In this chapter, the problem being examined and the theoretical rationale for the study are delineated. Next, the purpose and educational significance of the study is presented and the research questions outlined. Finally, terms used in the study are defined.

Statement of the Problem

Since the NAEYC first published the position statement on developmentally appropriate practices in 1987 (Bredekamp, 1987), the concept of developmental appropriateness has extensively influenced practice in the field of early childhood education, and the DAP statement has become one of the most influential documents guiding the field (Hart, Burts, & Charlesworth, 1997). Despite the popularity among practitioners and widespread acceptance of DAP guidelines, there have been few empirical studies investigating their effects on children's academic and social outcomes. Furthermore, existing empirical evidence from the limited number of research has been indecisive and controversial. Present studies on the effects of DAP guidelines are also limited in terms of methodological issues, such as the lack of appropriate analytical methods for nested data structures (Van Horn & Ramey, 2003), and large sample sizes.

Theoretical Framework/Rationale

Upon the arrival of the guidelines of the NAEYC, developmentally appropriate practice has had a major impact in the field of early childhood education. A number of controversies, confusion, and myths surrounded the debate on DAP (Mallory & New, 1994). The critiques were taken into account in revision of the position statement (Bredekamp, 1998).

Developmentally appropriate practice (Bredekamp, 1987; Bredekamp & Copple, 1997) is a term coined by the National Association for the Education of Young Children to denote the best possible practices, environments, and pedagogies for young children. The rationale behind DAP had its roots in certain philosophical and educational traditions, ranging from Jean Jacques Rousseau's romanticism to Piagetian constructivism and maturationist psychology. As Bredekamp and Rosegrant (1992) stated, developmentally appropriate practice is predominantly based on the theories of Piaget (1952), Erikson (1963), and Vygotsky (1978). Applications of Piaget's constructivist theory to developmentally appropriate practice are obvious in the documents published by NAEYC (see Bredekamp, 1987; Bredekamp & Copple, 1997; Bredekamp & Rosegrant, 1992; Bredekamp & Rosegrant, 1995).

Piaget's developmental theory has had an unprecedented influence on the field of early childhood education. Specifically, his theory has been used as a foundation for constructivist classrooms (Sheehan, 1979), for the application of mathematics education (Kamii, 1985, 1994) and for the design of early literacy programs (Ferreiro & Teberosky, 1982; Kamii, Manning, & Manning, 1991). Cognitive development was the major thrust of Piaget's theory (Piaget & Inhelder, 1969). While studies in cognition were primary, he studied moral development (Piaget, 1932), language development (Piaget, 1926) and even addressed emotional and affective development (Piaget, 1981).

Developmental theory of Piaget informed and guided one of the basic principles of DAP philosophy: “Human development research indicates that relatively stable, predictable sequences of growth and change that occur in children during the first 9 years of life” (Bredekamp & Copple, 1997, p. 10). It is claimed that knowledge of typical development of children within the age span grants a general framework to organize the learning environment, curriculum goals, and appropriate practices (Bredekamp & Copple, 1997). Although the stage theory has been highly criticized by many researchers (Flavell, 1982; Gelman & Baillargeon, 1983; Walsh, 1991), Smith (2002) argues that the stage criteria are not age-related, rather knowledge-related: “…developmental levels are levels of knowledge, not levels of knowers. Infant-level behavior in an adult is not a contradiction in Piaget’s model” (p.519). In other words, ages for each stage were considered as indicators, not criteria.

A Piagetian theoretical framework has also served as a basis for explaining how children construct their own understanding of the world around them as active learners.

NAEYC's position statement stipulates that children are active learners and they are "…actively engaged in constructing their own understanding from their experiences, and these understandings are mediated by and clearly linked to the sociocultural context. Young children actively learn from observing and participating with other children and adults…" (Bredecamp & Copple, 1997, p. 13). This statement illustrates Piaget's assertion that for a child to know and construct knowledge of the world, the child must act on objects and it is this action which provides knowledge of those objects (Piaget & Inhelder, 1969). Therefore, when active, the learner is not considered a vessel to be filled with facts. Affirming this belief means that teachers must provide meaningful opportunities for children to assume an active role in their own learning.

Another theory of constructivist development that affected the creation of developmentally appropriate practice was proposed by Vygotsky. Whereas Piaget's theory stresses the individual nature of the development, Vygotsky's theory is a sociocultural one, emphasizing socially constructed knowledge (Krogh, 1997). Vygotsky's sociocultural view underlines the importance of a child's interaction with others, adults or older peers, who mediate and support the child's learning experiences. In response to the criticism in the field that DAP does not meet the needs of a culturally and otherwise diverse populations (e.g., Graue, 1992; Jipson, 1991; Kessler, 1991; Lubeck, 1994; Ludlow & Berkeley, 1994), the second and latest version of the DAP statement by NAEYC acknowledges the role of children's social and cultural contexts to ensure that learning experiences are meaningful, relevant, and respectful for the participating children and their families (Bredekamp & Copple, 1997).

Unlike Piaget, who focused upon the child's independent construction of knowledge, Vygotsky's sociocultural view acknowledges the importance of a child's interaction with others, adults or more capable peers, who mediate and support the child's learning experiences (Vygotsky, 1978). His theory provided insight for the philosophy of DAP that learning occurs when children interact with both people and materials in their environments. Interactions between children and adults as well as other children facilitate children's mental manipulation and ownership of ideas. According to Vygotsky (1978), "Every function in the child's cultural development appears twice: first, on the social level, and later, on the individual level; first, between people (interpsychological) and then inside the child (intrapsychological). This applies equally to voluntary attention, to logical memory, and to the formation of concepts" (p.57). The social cognition learning model claims that culture is the primary determinant of individual development. Therefore, a child's learning development is affected by the surrounding culture, including the culture of family environment in which he or she is enmeshed.

A second essential tenet of Vygotsky's theory that contributed to the notion of DAP is the idea that the potential for cognitive development depends upon the "zone of proximal development" (ZPD) (Vygotsky, 1978): a level of development attained when children engage in social behavior. With appropriate adult help or in collaboration with more capable peers, children can often perform tasks that they are incapable of completing on their own (Vygotsky, 1978). According to Vygotsky (1978), an essential feature of learning is that it awakens a variety of internal developmental processes that are possible only when the child interacts with people in the environment and in cooperation with peers. This theory informs the relationship between the teacher and the child in DAP classrooms. Teachers provide a variety of activities and materials as well as encouraging the collaboration with other children; teachers adjust the level of difficulty of an activity and the level of help regarding the child's level of performance.

Regardless of the differences or similarities between the Piagetian and Vygotskian theories, the concept of developmentally appropriate practice proposes an interactive, constructivist view of learning (Bredekamp, 1987; Bredekamp & Rosegrant, 1992; Getswicki, 1999; Isenberg & Jalongo, 1997). Each of these theories explains the development of cognition as a process requiring active engagement in the experiential and social world for the young children. The principle that children construct their own knowledge through interactions with the social and physical environment is fundamental to this approach. Since the child is viewed as intrinsically motivated and self-directed, effective teaching can be achieved by considering children’s motivation to explore, experiment, and to make sense of their experience (Goffin & Wilson, 2001; Isenberg & Jalongo, 1997). Such a constructivist view of development calls for more direction from the child and less from the teacher. Due to this notion, DAP classrooms are generally referred as child-initiated (e.g., Marcon, 1992) or child-centered (e.g., Stipek, Feiler, Daniels, & Milburn, 1995) as opposed to the teacher-oriented or didactic (e.g., Stipek et al., 1995) classrooms where teacher is the main source of the curriculum. Although both the proponents and critics of DAP suggests that the practices must exist on a continuum (Bredekamp & Copple, 1997; Fowell & Lawton, 1993), the majority of the studies conducted on the effects of DAP share a dichotomous conceptualization of DAP.

This examination of DAP in two or three groups is due to the fact that NAEYC guidelines not only offer a framework for developmentally appropriate classrooms but also provide detailed examples of developmentally appropriate and inappropriate practices. In contrast with DAP, developmentally inappropriate practice (DIP) attempts to transfer knowledge through lecture and other whole-group activities (Hart, Burts, & Charlesworth, 1997). When comparing DAP and DIP classrooms in term of children's outcomes, a deep contrast is drawn between appropriate and inappropriate practices by its advocates (Elkind, 1986, 1987), who suggest that inappropriate classrooms should in fact hinder learning as opposed to the rich environment of the DAP classroom. If DAP's advocates are correct, it should be expected to see strong positive effects of DAP on children's learning and social development.

The effects of DAP on children's reading achievement are of particular interest because reading has been in the center of the recent accountability debate. Typically, DAP classrooms are expected to offer an environment that promotes both language acquisition and a positive sense of self, that allows children to take responsibility for their own learning, that is rich with print materials, and that provides more exposure to language (IRA & NAEYC, 1998) for children. Such recourses and experiences throughout the early childhood years, birth through age eight, are thought to positively affect the development of literacy. IRA and NAEYC believe that "given the range within which children typically master reading, even with exposure to print-rich environments and good teaching, a developmentally appropriate expectation is for most children to achieve beginning conventional reading (also called early reading) by age seven" (p. 8).

However, waiting to provide children with literacy experiences until they are school age can impede the level of reading they attain (Novick, 1996). Based on the evidence from the research that identifying letters (Adams, 1990; Stewenson & Newman, 1986), emergent spelling (Torgesen & Davis, 1996), and phonological awareness (Adams, 2001) are essential to later reading achievement, kindergarten classrooms that reflect a developmentally appropriate curriculum are “print-rich environments that provide opportunities and tools for children to see and use written language for a variety of purposes, with teachers drawing children’s attention to specific letters and words” (IRA & NAEYC, 1998, p. 9). DAP classrooms are also associated with positive growth in reading because they present “opportunities to engage in play that incorporates literacy tools, such as writing grocery lists in dramatic play, making signs in block building, and using icons and words in exploring a computer game” (p. 9). Previous research demonstrated that literacy–rich dramatic play is a robust predictor of later reading success (Dickinson & Tabors, 2001; Marrow, 1990; Neuman & Roskos, 1992). In all, investigating young children’s reading in developmentally appropriate and inappropriate classrooms is of significant interest to policy issues in the early childhood education.

Among other cognitive outcomes investigated in DAP research, creativity and divergent thinking are found to be fostered in developmentally appropriate classrooms (Hirsch-Pasek et al., 1990). Children also feel more confident in their abilities to solve intellectual problems (Kostelnik, Soderman, & Whiren, 2004). Although standardized tests after second grade in both DAP and traditional classrooms have shown little difference in general reading skills, children in DAP classrooms scored significantly higher in tests of vocabulary, reading comprehension, expressive language, and reading and writing mechanics in context (Nicholls et al., 1991).

Scores did not differ significantly in standardized tests of math in overall math skills, but the DAP children scored significantly higher in conceptual understanding and problem-solving skills than children in traditional classrooms. Nicholls and colleagues (1991) also found that DAP children display a better understanding of concepts and are more adept at generalizing computation skills across a variety of situations. In a controversial study, Hyson, Hirsh-Pasek, and Rescorla (1990) found no differences in academic achievement between DAP and teacher-initiated program preschoolers.

Greater enthusiasm for school, high involvement in the learning process and more cooperative behaviors are examples of positive social outcomes that DAP classrooms create (Kostelnik, Soderman, & Whiren, 2004). Researchers discovered that both children and teachers in developmentally inappropriate kindergartens demonstrated significantly more stress behavior (Burts, et al., 1990; Love, et al., 1992). Furthermore, children’s self-esteem (Curry & Johnson, 1990) and attitudes toward school (Hyson et al., 1990) tend to improve in developmentally appropriate classrooms.

Although the concept of DAP has been very popular and influential in the field of early childhood education, the number of empirical studies evaluating the effects of DAP have surprisingly been quite limited (Jones & Gullo, 1999; Van Horn & Ramey, 2003). It is far from being adequate to reach a conclusion about all aspects of the DAP. Furthermore, the results of the existing studies are weak and mixed (Van Horn & Ramey, 2003) to reach a consensus that differential effects of DAP are positive on children's academic achievement and growth. Of the studies that investigated the effects of DAP on cognitive domain, five studies (Burts, Hart, Charlesworth, DeWolf, Ray, Manuel et al., 1993; Dunn, Beach, & Kontos, 1994; Marcon, 1992, 1993, 1999) found positive effects, five studies reported mixed or no effects (Hirsh-Pasek et al., 1990; Huffman & Speer, 2000; Jones & Gullo, 1999; Stipek et al., 1998; Van Horn & Ramey, 2003), and one study reported negative effects (Stipek et al., 1995). Sample sizes of these studies have also been fairly limited; 37 to 307 children in 2 to 165 classrooms were included in various studies. Of those studies investigating the effects of DAP on academic achievement, only four were conducted in kindergarten, one in both kindergarten and first grade, and one in first grade. Assessment of DAP in those studies has also been highly criticized that some studies (e.g., Dunn et al., 1994) used an instrument that was not intended to measure DAP, rather more general constructs of child care quality (Hyson et al., 1990).

Along this vein, another methodological limitation of the previous studies on DAP arises. Unfortunately, most of the studies investigating the effects of DAP did not consider the nature of the nested data, that is to say that while the outcome variables, in most cases, are child level constructs such as children's cognitive or social outcomes, DAP is a classroom level variable. Children are nested within classrooms which in return are nested in schools. Van Horn and Ramey (2003) stated that most of the studies on DAP failed to use appropriate multilevel analytical methods for nested data structures. The most commonly observed analyses in DAP research, such as ANOVA, correlation, and regression, violate the assumption of independence (Tabachnick & Fidell, 1996). Because children in a classroom tend to share certain characteristics, individual observations are not fully independent. In the case of DAP research, children's DAP scores within the same classroom would be the same for each of them, meaning the each child's DAP score is dependent on the classroom. However, most analytic techniques entail independence of observations as a primary assumption for the analysis. Because this assumption is violated in the presence of hierarchical data, traditional data analyses produce standard errors that are too small, therefore, increase the likelihood of finding a significant difference (Type I error). Hox (1995, 1998) explains this phenomenon as the significance level of $p < .05$ of a test may in reality be as high as $p < .50$. This overestimation may have caused some of the significant effects of DAP reported in previous research. Hierarchical linear models (HLM) eliminate this type of bias by providing estimates of linear equations that explain outcomes for students in a group as a function of both the characteristics of the group and the characteristic of the individuals (Raudenbush & Bryk, 2002). With HLM, the nested structure of students within classrooms and classrooms within schools produces a different variance at each level for factors that are measured at that level.

Bryk, et. al. (1988) listed four advantages of HLM over regular linear models: (1) it can explain achievement and growth as a function of school level or classroom level characteristics, (2) it can model the effects of student characteristics, such as gender, race/ethnicity, or socioeconomic status, on achievement within schools or classrooms and then explain differences in these effects between schools or classrooms, (3) it can model the between and within group variance at the same time and produce more accurate estimates of student outcomes, and (4) it can produce better estimates of the predictors of student outcomes within schools and classrooms by using information about these relationships from other schools and classrooms.

In conclusion, the widespread appeal and acceptance of the DAP guidelines (Hart, Burts, & Charlesworth, 1997), growing criticism against some of the basic tenets of DAP (e.g., Kessler, 1991; Lubeck, 1994), limited number of the empirical studies (Jones & Gullo, 1999) and longitudinal studies (Stafford, van Rensburg, & Greene, 2000; Van Horn & Ramey, 2003) evaluating the effects of DAP, and the controversial, if not inconclusive, results of the previous research (Van Horn & Ramey, 2003) justifies the need for further research to find a broader consensus on the effects of DAP. This study was an attempt to examine the effects of DAP on children's reading achievement from kindergarten through third grade.

Purpose and Significance

The proposed study is an attempt to explore the effects of developmentally appropriate classroom practices on children's academic performance, namely reading scores, through the first four years of schooling. Subsequently, the effects of kindergarten DAP across levels of gender, SES, race/ethnicity, age, and school sector were also investigated. Considering the limitations of previous research methodology and the nature of the nested data structure, appropriate multilevel analyses were utilized to investigate the extent to which DAP in kindergarten is associated with children's reading achievement through third grade. Reading achievement is the outcome variable of interest which is measured at four time points in the study; two times in kindergarten, one time in first grade, and one time in third grade. The reading assessment included questions designed to measure basic skills (print familiarity, letter recognition, beginning and ending sounds, rhyming sounds, word recognition), vocabulary (receptive vocabulary), and comprehension (listening comprehension, reading passages in context).

The number of studies evaluating the effects of DAP has been very limited (Jones & Gullo, 1999; Van Horn & Ramey, 2003), despite the widespread popularity and acceptance of DAP guidelines. Furthermore, the results of the existing studies are weak and mixed (Van Horn & Ramey, 2003) to reach a consensus that differential effects of DAP are positive on children's academic achievement and growth. Whereas some studies reported positive effects of DAP (Burts et al., 1993; Dunn, Beach, & Kontos, 1994; Huffman & Speer, 2000; Marcon, 1992, 1993), mixed or no effects (Hirsh-Pasek et al., 1990; Jones & Gullo, 1999; Stipek et al., 1998), and negative effects (Stipek et al., 1995) in other studies were reported. Therefore, the current study is an attempt to shed light on the issue of whether the application of DAP in kindergarten improves children's reading achievement on two measures from kindergarten through third grade. Reading is assessed by two measures; first, a direct standardized measure, second, an indirect nonstandardized measure reported by teachers. The results from this study may assist the efforts of policy makers to create more effective schools.

Another striking finding when reviewing the research on DAP is that most studies examined the effects of DAP in isolation without considering its differential effects as a function of varying child, family, and teacher characteristics, such as SES, ethnicity, poverty, and teacher experience. Of the studies that examined the interaction between the effects of DAP and certain child variables, few found gender effects where boys did better in DAP classrooms and girls did better in DIP classrooms (e.g., Burts et al., 1992). However, two other studies found no interaction effects for gender (Marcon, 1992; Rawl & O'Tuel, 1981). Thus, the issue of whether some children benefit more from DAP remains unreciprocated and needs to be explored. Results of such analyses, may answer some of the major theoretical criticisms of the DAP guidelines that they reflect culturally biased values (Hsue & Aldridge; Smith, 1996). The current study is an attempt to expand the early findings and examine whether the initial and longitudinal effects of DAP are consistent across children from varying ethnicity, gender, poverty level, and home environment.

Finally, previous research has been limited with several methodological concerns. Most of the studies investigating the effects of DAP did not consider the nature of the nested data. That is, children are nested within classrooms, which in return are nested in schools. As Van Horn and Ramey (2003) stated, most of the studies on DAP failed to use appropriate multilevel analytical methods for nested data structures, resulting standard errors that are too small and increased likelihood of finding a significant difference (Type I error). Further, sample sizes of previous studies have also been fairly limited; 37 to 307 children in 2 to 165 classrooms were included in various studies. By employing an appropriate hierarchical linear modeling and drawing data from the Early Childhood Longitudinal Study-Kindergarten Class of 1998-99 (ECLS-K), the proposed study aims to overcome those two major methodological limitations of previous research. Thus, the current study aims to expand the body of knowledge regarding DAP and reading development in the primary grades as well as the methodological literature with nested data structure. The ECLS-K, conducted by the National Center for Educational Statistics (NCES), is a "multi-source, multi-method study that focuses on children's early school experiences beginning with kindergarten" (US Department of Education, 2000, p, 1-1). The ECLS-K gathers data from a nationally representative sample of children from kindergarten through fifth grade. A total of 21,260 children throughout the US are sampled in the study and assessed directly or indirectly by parent interview in the fall and/or spring of kindergarten. It is expected to better describe and understand the effects of DAP with this data because when used with appropriate sample weights (provided by NCES), results from the ECLS-K data are generalizable to the United States population of kindergarten children, teachers, and schools offering kindergarten programs in the 1998-1999 school year. Following research questions were addressed in this study:

1. Are there differences in the acquisition of reading knowledge and skills measured by the direct and indirect (ARS) assessment batteries across kindergarten, first grade and third grade by children's exposure to more developmentally appropriate classroom practices in kindergarten?
2. Do children benefit differently from the DAP in reading skills measured by the direct and indirect (ARS) assessment batteries by Age, Gender, Ethnicity, SES, and school sector?
3. Is there evidence for systematic change and individual variability in reading growth assessed through the direct and indirect (ARS) assessment batteries associated with teachers' use of developmentally appropriate practices over kindergarten year?
4. Does the effect of DAP on children's reading growth in kindergarten year vary by Age, Gender, Ethnicity, SES, and school sector?

Definitions of Terms

Child-Initiated Classroom Practices. These are observed in classrooms where daily schedules allow children to explore and develop sustained engagement with materials and peers, and support individualized transitions from one setting to another. They are believed to be responsive to each child's individual capabilities and needs. Curriculum includes rich opportunities for play and is responsive to the needs of the whole child—including cognitive and language, socioemotional, and physical development (Bredekamp & Copple, 1997)

Teacher-Directed Classroom Practices. Teacher-directed learning involves the teacher as a facilitator who models learning strategies and gives guided instruction. In this type of classrooms, the teacher initiates and leads the majority of activities while the entire class is involved.

Developmentally Appropriate Practices (DAP). DAP guidelines rely upon the principle that children construct their own knowledge through interactions with the social and physical environment. The definition of developmentally appropriate practice (Bredekamp, 1987; Bredekamp & Copple, 1997) argues that young children should be exposed to learning experiences that are in harmony with their current levels of development in order to aid and support their learning. The child is viewed as an active participant in the learning process; a participant who constructs meaning and knowledge through interaction with others, friends and family, materials and environment (Bredekamp, 1987; Bredekamp & Copple, 1997; Getswicki, 1999; Goffin & Wilson, 2001). Thus, more developmentally appropriate teaching practices are usually labeled as child-initiated or child-centered.

Developmentally Inappropriate Practices (DIP). These are the practices that are usually labeled as teacher-directed, didactic, or traditional practices.

Hierarchical Linear Modeling (HLM). HLM is a multilevel analysis of data that allows for the analyses of nested data at different levels simultaneously. HLM enables the testing of hypotheses about effects that occur within and between each level of the variables and about the interrelations between them (Raudenbush & Bryk, 1994).

Item Response Theory (IRT) Scores. Reading scores in this study were calculated using IRT procedures that considers the pattern of "right, wrong, and omitted responses to the items actually administered in a test and the difficulty, discriminating ability, and 'guess-ability' of each item to place each child on a continuous ability scale" (NCES, 2001). Scores reflect what the child would have achieved if all of the items had been administered. IRT scores provide several advantages to the researcher by compensating for the possibility of a low ability student guessing several hard items correctly, eliminating missing item responses, and providing indicators of student achievement on a common metric, even though the tests administered are not identical at each point (NCES, 2001).

Reading Growth. Reading achievement is the outcome variable of interest which is measured at six time points in the study; two times in kindergarten, two times in first grade, and two times in third grade. The reading assessment included questions designed to measure basic skills (print familiarity, letter recognition, beginning and ending sounds, rhyming sounds, word recognition), vocabulary (receptive vocabulary), and comprehension (listening comprehension, reading passages in context). The reading assessment contains five proficiency levels. These five levels reflect a progression of skills and knowledge; if a child had mastered one of the higher levels, he or she was very likely to have passed the items that comprised the earlier levels as well. These five levels were: (1) identifying upper- and lower-case letters of the alphabet by name; (2) identifying letters with sounds at the beginning of words; (3) identifying letters with sounds at the end of words; (4) recognizing common words by sight; and (5) reading words in context (US Department of Education, 2000). For the purposes of this study, a composite variable on an IRT based scale, that combines all those skills mentioned above, was utilized in a continuous form.

Sample Weights. Appropriate weights need to be used in any type of analysis with ECLS-K data to adjust the fact that data collection has a stratified design structure. The ECLS-K is not a simple random sample, that is, not all schools, teachers, and children had an equal probability of selection. For instance, private schools Asian/Pacific Islanders are oversampled to allow more powerful analyses using those variables. Because of this complex stratified sampling design, school-level, teacher-level, and child-level weights are provided and need to be used. Weights also allow researchers to adjust for differential nonresponses. Types of weight to be used in a specific analysis depend on the type of analysis (cross-sectional or longitudinal), level of analysis (child, teacher, or school), and the source of the data (child assessment, parent interview, teacher questionnaire etc.).

CHAPTER 2

LITERATURE REVIEW

This chapter begins with a brief history of developmentally appropriate practices (DAP). Basic principles and underlying theories of DAP are presented in the following section. The next section provides a detailed review of the studies examining the effects of DAP on cognitive and social outcomes. The final section presents a summary of this chapter.

History of Developmentally Appropriate Practices

The goal of early childhood education is to provide all children with experiences that allow them to develop and learn to their maximum potential. Growing numbers of children are cared for by people other than family and organizations in their early years (Getswicki, 1999). Due to society's increased recognition that the early years are significant and affect children's later experiences, the quality of childcare programs has been widely discussed in the early childhood education field by researchers and policy makers. In addition, national aspirations to remain competitive with other nations (Getswicki, 1999; Kostelnik, Soderman, & Whiren, 1999) have also influenced the field in a way that teachers and school administrators feel more pressure from the society in regard to the processes and outcomes of quality childcare. Elkind (1987) stated that this competitive urge resulted in reinforcement of the belief that earlier is better.

In schools, a continuing trend appears to be the *pushing down* of curriculum (Kostelnik, Soderman & Whiren, 2004) to the previous grade. That is, what used to be taught in first grade or later is now taught and assessed in kindergarten. Practices that are considered not to be appropriate for kindergarten children, such as whole-class teacher-directed instruction, grading, and using standardized test have been observed in kindergartens. Consequently, children face early failure and are frustrated with learning at early ages. Elkind (1981) and Gardner (1990) argued this unfortunate influence and warn that children are being hurried into functioning in ways that are not appropriate to their natural modes of learning.

Increasing concerns about the status of education, not only in early childhood education but also in the whole education system, initiated widespread calls for school reform (Bredekamp, Knuth, Kunesh, & Shulman, 2004) in the early 1980s. The question of *what to teach* guided critiques of existing curriculum content and approaches, resulting in many calls for change by such national organizations as the National Council of Teachers of Mathematics (1989), the National Council of Teachers of English (Lloyd-Jones & Lunsford, 1989), The National Association of State Boards of Education (1988), the Association for Supervision and Curriculum Development (1989), and others. In response to these arguments, the National Association for the Education of Young Children (NAEYC), the largest leading professional organization in the field of early childhood education, published a position paper (NAEYC, 1986) introducing the concept of developmentally appropriate practice. Bredekamp's (1987) statement, sponsored by NAEYC, followed this position paper. Bredekamp (1987) outlined both appropriate and inappropriate practices for programs serving children from birth through age eight. The DAP guidelines requires curriculum developers to take into account the following sources of curriculum (Spodek, 1991). Child development knowledge, individual characteristics of children, knowledge of various disciplines, knowledge of various cultures, parents' desires, and knowledge children need to function in the society.

The guidelines were introduced in order to boost the quality educational experiences for young children and foster professional identity and visibility for the field (Johnson & Johnson, 1992). NAEYC suggests that a high quality early childhood program should provide a "safe, nurturing environment that promotes the physical, social, emotional, and cognitive development of young children while responding to the needs of families" (Bredekamp, 1987, p.1). At the same time, the goal of DAP was to open up the curriculum and move away from the focus on academic skills and the formal teacher-directed drill and practice approach (Bredekamp & Rosegrant, 1992) to instruction.

Since the publication of the first guidelines (Bredekamp, 1987), DAP has been the subject of research examining the curricular approaches in early childhood education.

Although previous studies investigated the effects of various instructional approaches (e.g., Miller & Bizzell, 1983; Rawl & O'Tuel, 1981; Schweinhart, Weikart, & Larner, 1986), DAP guidelines have been accepted as the consensus starting point for the discussion of curriculum models in early childhood education (Jones & Gullo, 1999; Walsh, 1991) and provided a more constructed and organized perspective to assess the effects of those various approaches. Among the earlier studies, Montessori, High Scope, traditional, and cognitive developmental programs were regarded as more developmentally appropriate. On the other hand, DISTAR, DARCEE, Bereiter-Engelman, and action reading (Van Horn & Ramey, 2003) were considered as *developmentally inappropriate*.

The statement of the DAP guidelines (Bredekamp, 1987) fostered many efforts to create an instrument to measure the construct. Although DAP provided an opportunity to operationalize the instruction in early childhood classrooms and there is a widespread agreement on the definition of DAP, measuring to what extent a classroom is DAP still remains a controversial issue. Initial efforts started with the idea that teacher beliefs are the means to their practices and represent the level of developmental appropriateness in a classroom (Charlesworth, Hart, Burts, & Hernandez, 1991). This notion yielded instruments that assess teacher beliefs about DAP, rather than assessing the actual practice in the classroom. Later efforts for assessing DAP was also challenged with another controversy as whether it should be observed and rated by an outside observer or rated by teacher self-reports. Both approaches have their own disadvantages. For instance, teachers may alter their performance in the classroom when there is an outsider observes their instructional practices, and they may inflate the level of DAP in their classrooms because of DAP's popularity (Van Horn & Ramey, 2003). However, previous studies documented that teachers can be a valid and reliable source in judging not only their classroom practices (Burts et al., 1992) but also their students' intellectual, socioemotional, and behavioral development (e.g., Hopkins, George, & Williams, 1985; Meisels, Bickel, Nicholson, Xue, & Atkins-Burnett, 2001; Perry & Meisels, 1996). To date, there have been six instruments utilized in research to measure DAP: The Classroom Practices Inventory (CPI, Hirsh-Pasek et al., 1990), the Checklist for Rating Developmentally Appropriate Practice in Kindergarten Classrooms (Charlesworth et al., 1990), the Assessment Profile for Early Childhood Programs (Abbott-Shim & Sibley, 1992), A Developmentally Appropriate Practices Template (ADAPT, Gottlieb, 1997), the Early Childhood Environment Rating Scale (ECERS, Harms, Clifford, & Cryer, 1998), the Assessment of Practice in Early Elementary Classrooms (APEEC, Hemmeter, Maxwell, Ault, & Schuster, 2001, the Early Childhood Classroom Observation Measure (ECCOM, Stipek & Byler, 2004).The results of research studies that used these instruments are discussed in detail later in this chapter.

The Concept of Developmentally Appropriate Practice

The definition of developmentally appropriate practice (Bredekamp, 1987) argues that young children should be exposed to learning experiences that are in harmony with their current levels of development in order to aid and support their learning. DAP utilizes an approach to education of young children that focuses on the child as a developing human being and life-long learner. The child is viewed as an active participant in the learning process; a participant who constructs meaning and knowledge through interaction with others, friends and family, materials and environment (Bredekamp, 1987; Getswicki, 1999; Goffin & Wilson, 2001).

Piaget's theory of the construction of knowledge is seen as the basis for the cognitive-developmental approach (DeVries, 1997; Kamii & DeVries, 1993) with an emphasis on cognition. The teacher is an active facilitator in the system who helps the children make meaning of the various activities and interactions encountered throughout the day.

In the first version of NAEYC position statement (1986) developmentally appropriateness was viewed as consisting of two inter-related dimensions: Age appropriateness and individual appropriateness. This version of DAP was later updated in 1997 to include a consideration of social and cultural appropriateness (Getswicki, 1999; Goffin & Wilson, 2001) reflecting a more moderate mix of Piagetian and Vygotskian developmentalism than the first document.

Age appropriateness was defined in terms of theory and research that suggests universally predictable sequences of growth in children. As Bredekamp and Copple (1997) stated in the second version, age appropriateness stems from the "knowledge of age-related human characteristics that permits general predictions within an age range about what activities, materials, interactions, or experiences will be safe, healthy, interesting, achievable, and also challenging to children" (p.4). Activities, routines, and expectations are designed so that they accommodate and compliment the characteristics of children within a general age-range (Kostelnik, Soderman & Whiren, 2004). Knowledge of typical development of children within the age span served by the program provides a framework from which teachers prepare the learning environment and plan appropriate experiences.

Conversely, individual appropriateness was defined by the idea that each child is a unique character with an individual pattern and timing of growth (Bredekamp, 1987), as well as an individual personality and learning style (Kostelnik, Soderman & Whiren, 2004). Since the individual variation is seen as inevitable (Bredekamp & Copple, 1997), all possible variations must be considered in the design, application, and evaluation of activities, interactions, and expectations. The experiences provided should match the child's developing abilities, while also challenging the child's interest and understanding. Although these first two definitions appear contradictory, it is reconciled from the pedagogical perspective by the belief that each child will proceed through these stages at his/her own pace. The revised position statement stresses that age related data is to be used only for general predictions, without labeling the full range of individual levels of growth and ability as equivalent (Bredekamp & Copple, 1997). Hence, the children should be provided with learning experiences that match or suit their current levels of development. Teachers are required to make decisions in the classroom by combining their knowledge of child development with an understanding of the individual child to achieve desired and meaningful outcomes.

Children do not grow up in isolated areas, but rather, within families, neighborhoods, and communities. It is important that adults working with children have knowledge of the social and cultural contexts in which the children live in order to ensure that learning experiences are meaningful, relevant, and respectful for the participating children and their families. In the revised edition of the position statement, NAEYC suggests that the curriculum considers children within the context of their family, culture and community, past history, and present circumstances (Bredekamp & Copple, 1997). It was highlighted that early childhood professionals should recognize differences among children as well as characteristics they have in common in a cultural group. Developmentally appropriate practice approaches children as they are unique individuals. Relating and interacting with children and their parents are viewed as providers of information about this uniqueness. Parents' active involvement, both as resources of knowledge and as decision makers, is a necessity in deciding the individually appropriate practices for their children (Getswicki, 1999). Unlike the first version, group cultural differences are recognized as separate from individual differences. Cultural context is seen as an influence on behavioral expectations shared within a group. Practitioners of the field are to approach with sensitivity and respect to children and families from distinct cultural and linguistic groups.

These three dimensions as age, individual, and cultural appropriateness, represent the basic philosophy of developmentally appropriate practice and must be taken into account in planning programs. Getswicki (1999) informed teachers that as they make complex decisions for classrooms, decisions they make one year may be quite different from those will be made the next year. Developmentally appropriate practice is a philosophy that reflects child development knowledge in designing early childhood programs. It is critical in making developmentally appropriate decisions to be indicative of developmental theories and research.

Principles of Development Associated with DAP

Based on theories of Dewey, Vygotsky, Piaget, and Erikson, the concept of developmentally appropriately practice proposes an interactive, constructivist view of learning (Bredekamp, 1987; Bredekamp & Rosegrant, 1992; Getswicki, 1999; Isenberg & Jalongo, 1997). Each of these theories explains the development of cognition as a process requiring active engagement in the experiential and social world for the young children. Social development and emotional development are also focal points in DAP. The principle that children construct their own knowledge through interactions with the social and physical environment is key to this approach. Since the child is viewed as intrinsically motivated and self-directed, effective teaching can be achieved by considering children's motivation to explore, experiment, and to make sense of their experience (Goffin & Wilson, 2001; Isenberg & Jalongo, 1997).

A central issue in DAP is the role of play in the curriculum (Bredekamp & Copple, 1997). Children are viewed as using play as the vehicle through which the active engagement with the external world is utilized. Because adults don't depend on play to learn, they tend to dismiss it as a pleasant time spent without profit (Wolfgang & Wolfgang, 1999). In children, however, play is an essential part of a child's education. Sometimes called "children's work", play supports a child's development by providing the tools, equipment, and interpersonal experiences that help children grow. Through play, children acquire information, master activities, use concrete materials as symbols, organize previous learning, focus, solve problems, and develop creativity (Getswicki, 1999). Children's spontaneous play is a promoter of learning with opportunities for concrete, hands-on experiences. These experiences not only help children to master their environment but also enable them to develop the capacity for self regulation, abstract thought, imagination, and creativity. According to Vygotsky (1978), play and actively engaged acts lead development by providing a stage between the purely situational constraints of early childhood and adult thought, which is less context bound.

Another major premise of DAP is to make learning meaningful for the individual child, using practices which reflect both the age and individual needs of the child. A strong emphasis is given on learning to think critically, work cooperatively, and solve problems. Developmentally appropriate curriculum heavily depends on the theoretical foundation discussed above. Based on these theories, NAEYC offers 12 developmental and learning principles that must be considered in planning developmentally appropriate curriculum (Bredekamp & Copple, 1997):

1. Domains of development (physical, social, emotional, and cognitive) are inter-related and the child develops holistically.
2. Child development occurs in a relatively orderly and predictable sequence.
3. Development proceeds at varying rates within and among children.
4. Early experiences have both cumulative and delayed effects on children's later development.

5. Children's development proceeds in predictable directions toward greater complexity, organization, and internalization.

6. Development and learning occur in and are influenced by multiple social and cultural contexts.

7. Children construct their own knowledge in an attempt to gain an understanding of the world around them as active learners.

8. Children's learning is influenced by the interaction of biological maturation and the environment.

9. Play is an important vehicle for children's social, emotional, and cognitive development, as well as a reflection of their development.

10. Development advances when children have opportunities to practice newly acquired skills as well as when they experience a challenge just beyond the level of their present mastery.

11. Children's learning styles differ and they show what they know in various ways.

12. Children learn best when their physical and emotional needs are met and they feel safe and secure.

Practices Associated with DAP

In the position statement, the NAEYC outlined both developmentally appropriate and inappropriate practices for children from birth through age eight. Although the NAEYC lists many examples of appropriate and inappropriate practices that can be observed in education of young children, researchers in the field tend to agree on ten fundamental practices that characterize the DAP philosophy (Gullo, 1992; Hart, Burts, & Charlesworth, 1997; Kostelnik, Soderman, & Whiren, 2004; Miller, 1996). Practices that are evaluated as developmentally appropriate highly emphasize the importance of the following in early childhood classrooms:

1. The whole child. Children's emotional, social, cognitive, and physical needs should be addressed in designing the curriculum.

2. Individualizing the program to suit particular children. Developmentally appropriate practice encourages the use of varied instructional strategies to meet the various learning needs and interests of children in the group.
3. Importance of child-initiated activity. Developmentally appropriate practice encourages a mixture of teacher-directed and child-directed activities in a setting where children are viewed as active decision makers and responsible from their own learning.
4. Play as a vehicle for learning. Both outdoor and indoor learning is valued as a medium for learning.
5. Flexible, stimulating classroom environments. A balance of direct and indirect instruction should promote the child's learning as needed.
6. Integrated curriculum. It combines many subject areas (e.g., math, science, and reading) into an interconnected unit of study that is meaningful to students. The integrated activities are often related to real life.
7. Learning by doing. Developmentally appropriate programs promote children's active learning experiences in which children manipulate real objects and learn through hands-on, direct activities. Learning centers serve as means of providing active learning experiences.
8. Providing choices to children. Teachers are to provide a wide variety of ways to learn, so that children with various learning styles are able to develop their own capabilities and are enabled to view learning in new ways.
9. Continuous assessment of individual children and the program as a whole. Standardized tests are de-emphasized, while a variety of formal and informal assessment techniques are valued to determine the needs of each child.
10. Partnership with parents. Parents serve as both decision makers and sources of information about each child. The partnership is seen essential to achieve programs that are responsive to the needs o individuals.

According to the NAEYC, developmentally appropriate practice provides children with opportunities to learn and practice newly acquired skills. Regarding Vygotsky's theory of the zone of proximal development, DAP offers challenges just beyond the level of their present mastery and it takes place "in the context of a community where children are safe and valued, where their physical needs are met, and where they feel psychologically secure" (Bredekamp & Copple 1997, pp. 14-15). With appropriate adult help or in collaboration with more capable peers, children can often perform tasks that they are incapable of completing on their own (Vygotsky, 1978). According to Vygotsky (1978), an essential feature of learning is that it awakens a variety of internal developmental processes that are possible only when the child interacts with people in the environment and in cooperation with peers. This theory informs the relationship between the teacher and the child in DAP classrooms. Teachers provide a variety of activities and materials as well as encouraging the collaboration with other children; teachers adjust the level of difficulty of an activity and the level of help regarding the child's level of performance.

The classroom environment provided in developmentally appropriate classrooms are of particular interest in investigating young children's reading growth. Typically, DAP classrooms are expected to offer an environment that promotes both language acquisition and a positive sense of self, that allows children to take responsibility for their own learning, that is rich with print materials, and that provides more exposure to language (IRA & NAEYC, 1998) for children. Such recourses and experiences throughout the early childhood years, birth through age eight, are thought to positively affect the development of literacy. IRA and NAEYC believe that "given the range within which children typically master reading, even with exposure to print-rich environments and good teaching, a developmentally appropriate expectation is for most children to achieve beginning conventional reading (also called early reading) by age seven" (p. 8).

However, waiting to provide children with literacy experiences until they are school age can impede the level of reading they attain (Novick, 1996). Based on the evidence from the research that identifying letters (Adams, 1990; Stewenson & Newman, 1986), emergent spelling (Torgesen & Davis, 1996), and phonological awareness (Adams, 2001) are essential to later reading achievement, kindergarten classrooms that reflect a developmentally appropriate curriculum are "print-rich environments that provide opportunities and tools for children to see and use written language for a variety of purposes, with teachers drawing children's attention to specific letters and words"

(IRA & NAEYC, 1998, p. 9). DAP classrooms are also associated with positive growth in reading because they present "opportunities to engage in play that incorporates literacy tools, such as writing grocery lists in dramatic play, making signs in block building, and using icons and words in exploring a computer game" (p. 9). Previous research demonstrated that literacy–rich dramatic play is a robust predictor of later reading success (Dickinson & Tabors, 2001; Marrow, 1990; Neuman & Roskos, 1992). In all, investigating young children's reading in developmentally appropriate and inappropriate classrooms is of significant interest to policy issues in the early childhood education.

Reading in Early Years

Reading is the fundamental skill in a child's life upon which the success in other learning domains depends. Mastery of cognitive and socio-emotional skills and knowledge in later grade is, to a great extent, related to a child's ability to comprehend reading basics in early years. Therefore, studying early reading and language development has an invaluable importance to shape policy decisions. Research has demonstrated the importance of early years in reading as children who read poorly in first and second grade tend to remain poor readers in later grades (Blachman, 2000; Snow, Burns, & Griffin, 1998).

Lyon (1999) stated that a large number of students fail to acquire basic reading skills in early elementary grades. Of those children who struggle to achieve basic levels of reading by third grade, 75% are expected to be poor readers at the end of high school (Francis, Shaywitz, Stuebing, Shaywitz, & Fletcher, 1996; Lyon, 1998). Consequently, students who learn to read and acquire basic language skills in kindergarten attain higher levels of reading at the end of high school (Hanson & Ferrel, 1995). Research extensively illustrated that children who are poor readers at the end of first grade generally do not obtain average-level reading skills by the end of elementary school (Francis et al., 1996; Juel, 1988; Torgesen and Burgess, 1998). Children's abilities to comprehend and learn new concepts are influenced by their difficulties with word recognition and fluency (Juel, 1988; Torgesen, Wagner, & Rashotte, 1994). Furthermore, children who have difficulties in reading tend to have negative attitudes toward reading, fewer opportunities for vocabulary growth, and less practice in reading (Stanovich, 1986) than their peers who attain desired levels of reading.

Given the importance of beginning reading skills in early years, a debate over various approaches to teach those skills has shaped the literature and research efforts in the past two decades. Whereas some researchers supported an approach called *phonics*, others promoted the *whole-language* approach in reading instruction. Phonics method is typically regarded as a decoding process, in which children associate letters with sound and decode whole words (Ehri, 1992; Perfetti, 1992; Reitsma, 1983; Share, 1999). Direct instruction, practice, and reinforcement follow each other until the child masters a basic text, then a more complex text. Because of the influence of behaviorist theory, learning the alphabetical code and word recognition are often taught as isolated skills, rather than in the context of reading or writing. Crawford (1995) explains this position as "the act of reading (and writing) can be broken down into a series of isolated skills, which can be arranged into a hierarchy, taught directly, and then brought back to the whole" (p. 78).

In contrast, the whole language method emphasizes the meaning of whole words in the context sentences and paragraphs (Clay, 1991; Goodman, 1986). Actually reading and writing in practice is seen essential to learning to read and write, just as children learn to talk in conversations (Baron, 1990). Researchers and professional organizations focused their efforts to find the essential components of a comprehensive reading program from both schools of thought that provide a balance of best experiences. The report by the National Research Council (Snow et al., 1998) documented the importance of appropriate high-quality reading instruction in grades K-3 to prevent reading difficulties. A recent report of the National Reading Panel (2000) acknowledged five critical components of early reading instruction as: phonemic awareness, phonemic decoding skills (phonics), fluency in word recognition and text processing, reading comprehension strategies, and vocabulary. These elements are viewed essential to foster children's effective learning of basic early reading skills.

Longitudinal studies of children's literacy have scrutinized the cognitive skills that fortify reading and spelling development. Bradley and Bryant (1983) conducted one of the most influential of these studies (Snowling, 2002) to examine the relationship between early phonological skills and later reading achievement in some 400 children from 4 to 8 years of age. Children were administered three phonological tests at the beginning of the study and two of those tasks, rime oddity and alliteration oddity, were associated with the later reading achievement. A strong relationship between the children's phonological awareness assessed on those two tasks at 4 years and their reading and spelling skills at 8, was demonstrated in Bradley and Bryant's (1983) study, even when the substantial effects of IQ, memory, and social class were controlled for.

Another key study was conducted in Denmark where, at the time, children were not exposed to reading instruction before the age of 7 years. Lundberg and colleagues (Lundberg, Olofsson, & Wall, 1980) replicated the strong relationship between phonological awareness prior to literacy instruction and later reading achievement and provided evidence for the universality of these findings. This relationship has also been confirmed by a large number of studies (e.g., Ellis & Large, 1987; Tunmer & Nesdale, 1985).

Research on DAP

Although a growing body of research from diverse fields such as developmental psychology, cultural anthropology, linguistics, early childhood education, and brain research have been conducted about DAP or child-initiated approach, it is far from being adequate to reach a conclusion about all aspects of the DAP. Upon the arrival of the guidelines of the NAEYC, a number of controversies, confusion, and myths surrounded DAP (Mallory & New, 1994). The critiques were taken into account in revision of the position statement (Bredekamp, 1998). Since the publication of DAP statement, the construct of developmental appropriateness has been investigated by educational researchers. Several studies have examined the effects of DAP on cognitive outcomes (e.g., Burts et al., 1993; Dunn, Beach, & Kontos, 1994; Hirsh-Pasek, Hyson, & Rescorla, 1990; Huffman & Speer, 2000; Jones & Gullo, 1999), whereas others studied social and emotional outcomes (e.g., Burts, Hart, & Kirk, 1990; Jones & Gullo, 1999; Marcon, 1992; Stipek et al, 1995) affected by the construct. Despite the growing number of empirical research on the topic, evidently, there are still potential questions to be explored regarding the effectiveness of DAP in the context of early childhood education. The lack of research on the effects of DAP in relation with other student and classroom variables (Jones & Gullo, 1999) and the inadequacy of previous research methodology

(Van Horn & Ramey, 2003) on DAP have been recognized in the field. Particularly, the need for investigating longitudinal effects of kindergarten DAP on children's academic achievement has been addressed by researchers in the literature (Burts et al., 1993; Jones & Gullo, 1999; Van Horn & Ramey, 2003). The results of published empirical studies examining the effects of DAP on cognitive and social student achievement is presented in the next section. Studies reviewed under each domain are summarized in Table 1 for a more detailed examination including such features as the sample, instruments used, methodology, and findings.

Research on Cognitive Outcomes

Although the concept of DAP has been very popular and influential in the field of early childhood education, the overall number of empirical studies evaluating the effects of DAP have surprisingly been quite limited (Jones & Gullo, 1999; Van Horn, Karlin, Ramey, Aldridge, & Snyder, 2005; Van Horn & Ramey, 2003). It is far from being adequate to reach a conclusion about all aspects of the DAP. Furthermore, the results of the existing studies are weak and mixed (Van Horn & Ramey, 2003) to reach a consensus that differential effects of DAP are positive on children's academic achievement and growth. Of the 11 published studies that examined the effects of DAP on cognitive domain, five studies revealed positive effects (Burts et al., 1993; Dunn, Beach, & Kontos, 1994; Marcon, 1992, 1993, 1999), five studies reported mixed or no effects (Hirsh-Pasek et al., 1990; Huffman & Speer, 2000; Jones & Gullo, 1999; Stipek et al., 1998; Van Horn & Ramey, 2003), and one study reported negative effects (Stipek et al., 1995). Sample sizes of these studies have generally been fairly limited; 37 to 721 children in 2 to 165 classrooms were included in various studies. The most sophisticated study by Van Horn and Ramey (2003) provides a rare example of adequate sample size and statistical analysis in DAP research; the sample included 4764 children in 1537 classrooms and a multilevel statistical analysis was employed. Of those studies investigating the effects of DAP on academic achievement, only four were conducted in kindergarten, one in both kindergarten and first grade, and one in first grade. Assessment of DAP in those studies has also been highly criticized that some studies (e.g., Dunn et al., 1994) used an instrument (ECERS) that was not intended to measure DAP, rather more general constructs of child care quality (Hyson et al., 1990).

Two studies by Marcon (1992, 1999), compared preschool children in child initiated (CI), academically directed (AD) and middle-of-the-road (M) classrooms. Children in both the model CI (DAP) and the model AD (DIP) classrooms outscored children in the model M classrooms on the Vineland Adaptive Behavior Scales scores and district's progress report scores. Although the magnitude of the difference was not as large, model CI children were significantly better than model AD children on Vineland subscales of receptive and expressive language skills, personal, interpersonal, and gross motor skills. Model CI children also performed better than model AD children on all subject areas of the report card. However, model AD children were better than model CI and M children in written language and play and leisure time skills measured by the Vineland instrument. Another study by Marcon (1993) compared a sample of predominantly (94%) African American 307 kindergartners in classrooms in which either socioemotional growth (DAP) or academic preparation (DIP) was emphasized. Results indicated a main effect for the kindergarten model on report card scores and an interaction effect of kindergarten model and gender on Vineland scores. Children from classrooms in which teachers emphasized socioemotional development received better grades in science, physical science, and social science. Furthermore, boys had better adaptive skills in socioemotional classrooms, whereas girls had better adaptive skills in academic classrooms.

Using a sample of 166 kindergarten children from a previous study (Burts et al., 1992), Burts et al. (1993) reported that first graders from more DAP kindergartens received better grades than children from less DAP kindergarten classrooms in reading, math, science, and social studies. The effect of DAP was differential on children from low SES and high SES families. Low SES children had better overall GPAs in more DAP classrooms, whereas high SES children performed better in less DAP classrooms. Two more studies found positive effects of DAP on cognitive outcomes of learning. Stipek et al., (1998) and Dunn et al., (1994) found that DAP classrooms were associated with better academic outcomes. In all, results of these six studies proposes a positive effect of

DAP on various cognitive outcomes. However, the effects of DAP were not consistent through other research.

Another group of studies reported mixed or negative effects of DAP that put the efforts and finances spent on DAP into question. In a study, a sample of predominantly African-American (76%) and minority (99%) 113 kindergarten and first grade children were examined in moderately DAP and low DAP classrooms (Huffman & Speer, 2000). Students in the former classrooms scored better in letter-word identification and applied problems, but not in math calculations. However, this was an interaction effect mediated by the semester. The main effects of DAP were not significant. In another study of 90 preschool children, Hirsh-Pasek et al. found no significant effects of DAP over DIP on several measures of cognitive achievement. Stipek et al., (1998) compared 228 children in less basic skills oriented (DAP) and more basic skills oriented (DIP) classrooms. Results from this study were controversial as children from less basic skills oriented classrooms performed better in preschool but worse in kindergarten on math and reading scores. Jones and Gullo (1999) examined the effects of DAP with a sample of 293 first graders. They reported that there were no differences between DAP and DIP classrooms on five measures of mathematical achievement and three measures of language achievement. Contrary to the results from Marcon's (1992, 1999) studies, children whose teachers adopted neither DAP nor DIP (average) performed better in two measures of math than students from other two classrooms.

Table 1
Summary of Research Examining the Effects of DAP

Author(s)	Sample	Outcome Measure(s)	DAP Measure(s)	Type of Analyses	Results
Burts, Hart, Charlesworth, & Kirk (1990)	37 children in 2 classrooms, 5-6 year old kindergartners	Observer rated Child Stress Behavior Instrument (CCSBI) including four factors: passive, self with self, self with others, self with object; Frequency of time spent in centers and activities	Teacher reported Teacher Beliefs Scale (TBS) and Instructional Activities Scale (IAS), observer rated Checklist for Rating Developmentally Appropriate Practice	MANOVA	Children in DIP classrooms exhibited more overall stress behaviors. More stress behaviors were found in whole-group and workbook activities in DIP classrooms; whereas more stress was found in center and transition activities in DAP classrooms. DAP classrooms spent more time in centers, group story, and transition activities, DIP classrooms spent more time in whole-group and workbook/worksheet activities.
Hirsh-Pasek, Hyson, & Rescorla (1990)	90 pre-K children in 11 classrooms, a subset of 56 children followed-up in K	Academic Skills Inventory, PASS, Ravens Coloured Progressive Matrices, Torrance Test of Preschool Creative Thinking, emotional well-being	CPI	*t*-test, correlation	No significant differences were found on any outcome variable between high academic schools and low academic schools. Only difference that was close to significance (p=.10) was observed as higher test anxiety in high academic schools.
Burts, Hart, Charlesworth, Fleege, Mosley, & Thomasson (1992)	204 kindergarten children in 12 classrooms	Observer rated Child Stress Behavior Instrument (CCSBI) including four factors: passive, self with self, self with others, self with object	Teacher reported Teacher Beliefs Scale (TBS) and Instructional Activities Scale (IAS), observer rated Checklist for Rating Developmentally Appropriate Practice in Kindergarten Classrooms	ANOVA MANOVA	Children in DIP classrooms exhibited more overall stress behavior. Boys in DIP classrooms showed more stress than boys in DAP classrooms. No difference for girls. In DIP classrooms, Black children had more stress than White children during transition, waiting, and workbook/worksheet activities. No differences were found in DAP classrooms by race or gender.

Table 1 (continued)

Author(s)	Sample	Outcome Measure(s)	DAP Measure(s)	Type of Analyses	Results
Marcon (1992)	295 pre-K children aged 4 and 5 in 39 schools.	Vineland Adaptive Behavior Scales including four factors: communication, daily living skills, socialization, and motor development; GPA derived from district's Early Childhood Progress Report	Pre-K Survey of Beliefs and Practices	Hierarchical cluster analysis to group teachers; ANOVA MANOVA MANCOVA	Children in middle-of-the-road classrooms scored lower than children in child-initiated (CI) and academically directed (AD) classrooms. No differences were found between CI and AD classrooms.
Burts, Hart, Charlesworth, DeWolf, Ray, Manuel, & Fleege (1993)	166 first grade children followed-up from kindergarten	GPA derived from first-grade report cards	TBS and IAS	ANOVA MANOVA	High SES children did not differ in DAP classrooms. Low SES children had better overall GPA in DAP classrooms. In DIP classrooms, high SES children had better overall GPA than low SES children. Children in DAP classrooms had better grades in reading, math, science, and social studies than children in DIP.
Marcon (1993)	307 low income Head Start and pre-K children in 86 classrooms at 56 schools	Vineland Adaptive Behavior Scales including four factors: communication, daily living skills, socialization, and motor development; GPA derived from district's Early Childhood Progress Report	Pre-K Survey of Beliefs and Practices	Hierarchical cluster analysis to group teachers; ANCOVA MANCOVA	Children in socioemotional and academic classrooms did not differ in Vineland scores. Children in socioemotional classrooms had better school achievement than children in academic classrooms. In academic classrooms, girls had better adaptive skills than boys. Boys benefited more in socioemotional classrooms.

Table 1 (continued)

Author(s)	Sample	Outcome Measure(s)	DAP Measure(s)	Type of Analyses	Results
Dunn, Beach, & Kontos (1994)	60 pre-K children in 30 classrooms	Classroom Behavior Inventory, Preschool Inventory-Revised Edition	ECERS	Correlation, regression	Children's receptive language was better in higher quality literacy environments with more DAP activities. DAP had no effects on PSI achievement scores.
Stipek, Feiller, Daniels, & Milburn (1995)	227 pre-K and K children in 32 classrooms	Woodcock-Johnson achievement test, Peabody Individual Achievement Test, various direct measures of children's motivation (e.g., enjoyment of school, perceptions of ability, dependence, and anxiety)	An observation measure adapted from a previous study by Stipek et al., (1992), ECERS, CPI	ANCOVA MANCOVA	Children in didactic classrooms outperformed children in child-centered classrooms in letters/reading achievement, but displayed more dependency, more anxiety, less pride in accomplishment, less preference for challenge, lower expectations for success, and lower perceptions of ability.
Hart et al. (1998)	102 predominantly (90%) white pre-K children	Classroom Child Activity Instrument, Observer rated Child Stress Behavior Instrument (CCSBI) including four factors: passive, self with self, self with others, self with object;	Teacher reported Teacher Beliefs Scale (TBS) and Instructional Activities Scale (IAS), observer rated Checklist for Rating Developmentally Appropriate Practice	MANOVA	Children in DIP classes spent more time on waiting, workbook, and TV watching, while DAP children spent more time on transition, music, group story, and center activities. DIP children exhibited more overall stress behavior. Interaction effect for SES and DAP reported. In DIP classes, low SES preschoolers had more stress behavior than high SES children; same effect were not found in DAP classes.

Table 1 (continued)

Author(s)	Sample	Outcome Measure(s)	DAP Measure(s)	Type of Analyses	Results
Stipek et al., (1998)	228 pre-K and K children in 42 classrooms, followed in the next year	Woodcock-Johnson achievement test, Peabody Individual Achievement Test, McCarthy General Cognitive Competence test, various direct measures of children's motivation	An observation measure adapted from a previous study by Stipek et al., (1992), ECERS, CPI	ANOVA ANCOVA MANOVA	Children in less basic skills oriented (DAP) classrooms generally performed better in preschool. Kindergartners in basic skills oriented (DIP) classes performed better in cognitive outcomes. Mixed results were found on the continuation of DAP from preschool to kindergarten and 1st grade.
Jambunathan, Burts, & Pierce (1999)	91 preschool children aged 3 to 5 in 11 clasrooms	Pictorial Scale of Perceived Competence and Social Acceptance	Checklist for Rating Developmentally Appropriate Practice in Early Childhood Classrooms	Step-wise regression	Use of DAP explained 15% of variance in peer acceptance scores in preschool.
Jones & Gullo (1999)	293 first grade children in 13 classrooms	Integrated Assessment System, Social Skills Rating System	TBS and IAS	ANOVA	Children in Average (neither DAP or DIP) classes had better math. No differences were found on reading achievement. DAP children had higher scores on social skills than children in Average and DIP classes.
Marcon (1999)	721 four-year-olds in 114 classrooms at 52 schools	Vineland Adaptive Behavior Scales including four factors: communication, daily living skills, socialization, and motor development; GPA derived from district's Early Childhood Progress Report	Pre-K Survey of Beliefs and Practices	ANCOVA MANCOVA	Children in middle-of-the-road (M) classrooms scored lower than children in child-initiated (CI) and academically directed (AD) classrooms on most developmental skills. No differences were found between CI and AD classrooms. CI children performed better in school than AD and M children on all subject areas.

Table 1 (continued)

Author(s)	Sample	Outcome Measure(s)	DAP Measure(s)	Type of Analyses	Results
Ruckman, Burts, & Pierce (1999)	50 low SES African American 1st grade students in 4 classrooms	Classroom Child Stress Behavior Instrument s	Researchers rated activities in computer labs based on NAEYC guideline	Chi-square	Children in more DAP classes exhibited less stress behavior than children in less DAP classes. No significant gender effects were found.
Huffman & Speer (2000)	113 low income minority K and 1st grade children	Woodcock-Johnson Psycho-Educational Battery-Revised	Assessment Profile for Early Childhood Programs	MANOVA	No main effects of DAP were found. Children in Moderate DAP classes performed better in spring semesters on letter-word identification and applied problems than children in Low DAP classes. No significant interaction effects were found on math calculations mediated by semester.
Van Horn & Ramey (2003)	Former Head Start students, 4764 first graders, 2,690 second graders, and 1,569 third graders in 859 to 1,537 classrooms	Peabody Picture Vocabulary Test-Revised, Woodcock-Johnson achievement test	ADAPT	Four-level growth curve analysis	No significant effects of DAP were found. Interactions effects by gender, ethnicity, and poverty level were also found to be nonsignificant.

In the most recent study of DAP, Van Horn and Ramey (2003) examined 1,569 to 4,764 children from first through third grades. Using an appropriate multilevel longitudinal design and a large sample size, Van Horn and Ramey did not find any significant effect of DAP in any grade level and growth over time on any of the outcome measures. However, this study used a sample of predominantly (68% to 72%) former Head Start children.

Finally, the study by Stipek et al., (1995) compared 227 preschool and kindergarten children in 32 classrooms. Children in didactic (DIP) classrooms that stressed basic skills performed better on a letters/reading achievement test than children from child-centered classrooms. This effect was not served on a numbers achievement test, though.

To sum up, the findings of the available studies that examined the effects of DAP on cognitive measures of achievement, are mixed and controversial. Whereas five studies reported positive effects of DAP, five other studies revealed mixed or no effects and one study found negative effects of DAP. Furthermore, some studies reported no main effects of DAP but some interaction effects mediated by child characteristics as gender and SES. However, the effects of DAP in relation with gender, SES, and race/ethnicity was only scrutinized in one study together. The results of the available studies are mixed and inconclusive about the differential effects of DAP; some group of children benefited more from DAP, while others had a disadvantage in DAP classrooms. Results were not generalizable to the population of children because some of the studies used a sample of children that mostly belonged to one race or socioeconomic status (e.g., Huffman & Speer, 2000; Ruckman et al.). Furthermore, only one study investigated the effects of kindergarten DAP on a later grade, but that study was limited with the performance at the first grade. The effects of kindergarten DAP beyond first grade remains yet unexplored. Finally, all the studies, but the one conducted by Van Horn and Ramey (2003), were limited by either a small sample size or by a lack of appropriate multilevel research design that would fit the nested nature of the data. That is, while the outcome variables, in most cases, are child level constructs such as children's cognitive or social outcomes, DAP is a classroom level variable. Children are nested within classrooms, which in return are nested in schools. Ten of the eleven available studies used a research design that did not take the nested nature of the data into account, such as regression and MANOVA. Failure to utilize a multilevel nested design the presence of hierarchical data results in producing standard errors that are too small, therefore, increases the likelihood of finding a significant difference (Type I error) (Hox, 1995, 1998).

Research on Psychosocial Outcomes

Although the effects of DAP on children's cognitive learning are inconclusive and mixed, research provided more consistent findings relating DAP to better social skills of young children. Decreased numbers of overall stress, dependency to others, and anxiety as well as increased expectancy for success, higher perceptions of ability, and more pride in accomplishment are some of the positive effects that are associated with DAP. It is argued that even if there is no cognitive effects, positive effects of DAP on psychosocial outcomes supports endorsement of a more developmentally appropriate curriculum model alone (Stipek et al., 1995).

As it was the case with cognitive outcome research, the overall number of empirical studies evaluating the effects of DAP on psychosocial outcomes has been limited (Jones & Gullo, 1999; Van Horn et al., 2005; Van Horn & Ramey, 2003), despite the popularity of the concept of DAP. Of the 12 published studies that examined the effects of DAP on psychosocial domain, six studies (Burts et al., 1990, 1992; Hart et al., 1998; Jones & Gullo, 1999; Ruckman, Burts, & Pierce, 1999; Stipek et al., 1995) reported positive effects and another six studies reported mixed, small positive or no effects (Hirsh-Pasek et al., 1990; Jambunathan, Burts, & Pierce, 1999; Marcon, 1992, 1993, 1999; Stipek et al., 1998). These studies were also limited by some design issues as small sample sizes and the lack of multilevel statistical analysis. Of those studies investigating the effects of DAP on psychosocial achievement, only six were in preschool, three in kindergarten only, one in both kindergarten and first grade, and two in first grade. The psychosocial effects of DAP beyond first grade remains yet unexplored.

Hart et al., (1998) compared children's stress behaviors in DAP and DIP classrooms. Preschoolers in DIP classrooms displayed twice as much stress behaviors compared to the children in DAP classrooms. In addition, a significant interaction effect with SES indicated that low SES children in DIP classrooms exhibited more stress behaviors than did high SES children in DIP classrooms. In another study, Stipek et al. (1995) reported that children in didactic classrooms that emphasized teaching basic skills scored lower on most of the measures of motivation. That is, children in didactic classrooms displayed more dependency, more anxiety, less pride in accomplishment, less preference for challenge, lower expectations for success, and lower perceptions of ability, compared to children in child-centered classrooms. Two other studies by Burts and colleagues (Burts et al., 1990, 1992) linked DAP to positive psychosocial measures in kindergarten. In these two studies, children from DIP classrooms displayed more overall stress behaviors than did their peers in DAP classrooms. Furthermore, two interaction effects with gender and SES were found as boys in DIP classrooms showed more stress than boys in DAP classrooms. No differences were observed for girls. In DIP classrooms, African America children had more stress than Caucasian children during transition, waiting, and workbook/worksheet activities. No differences were found in DAP classrooms by race or gender.

Two more recent studies examined the social effects of DAP in first grade. Jones and Gullo (1999) reported that children in DAP classrooms had higher scores on social skills than children in Average and DIP classes, measured by the Social Skills Rating System. There were no differential effects in DAP classrooms. The second study (Ruckman et al., 1999) of 50 low income African American children also indicated a similar positive effect as children in less DAP classrooms exhibited more stress behaviors than children in DAP classrooms. Further analyses for interaction effects did not reveal any significant gender effects associated with DAP.

Another group of six studies reported small or no effects of DAP on children's socioemotional outcome measures. The first study conducted by Hirsh-Pasek et al., (1990) found no significant differences on any outcome variable between high academic schools (DIP) and low academic (DAP) schools. The only difference that was close to significant (p=.10) was observed as higher test anxiety in high academic schools compared to lo academic schools. Three studies by Marcon (1992, 1993, 1999) compared children from child-initiated (DAP), middle-of-the-road (M) and academically directed (DIP) classrooms. In general, children in the M classrooms consistently scored lower than other groups on all measures of social skills. This finding is controversial and challenging to the recent statements of NAEYC (Bredekamp & Copple, 1997) that DAP classrooms should maintain a balance of more developmentally appropriate activities and inappropriate activities. However, no differences between DAP and DIP classrooms were found. In another study, Stipek et al., (1998) presented that children in more basic skills oriented (DIP) classrooms were observed engaging in more prosocial behaviors with peer, had higher expectations for performance, and enjoyed school activities more than did children in less basic skills oriented (DAP) children. Children in DIP classrooms, on the other hand, also displayed more noncompliant behavior, were disciplined more, and were more dependant on others for approval than children in DAP classrooms. The mixed results on several measures of social skills in this study were in accord with another study by Jambunathan et al., (1999). The findings from this study revealed that DAP was a significant predictor of peer acceptance domain of the preschoolers' perception of self-competence. However, DAP was not a significant predictor of other three measures of self-competence; cognitive competence, physical competence, and maternal acceptance.

Results of the studies that examined the effects of DAP on psychosocial measures of achievement, are more uniformed than results from studies investigating cognitive measures. Of the 12 published studies, six reported positive effects of DAP, four reported no or small effects, and only two reported mixed results favoring both DAP and DIP on varying measures of social skills. In general, findings on the psychosocial outcomes are judged to be in favor of DAP (Stipek et al., 1995; Van Horn et al., 2005). Also, results were not generalizable to the population of children because some of the studies used a sample of children that mostly belonged to one race or socioeconomic status (e.g., Hart et al., 1998). Furthermore, all the studies suffered from psychometric issues such as a small sample size and a lack of appropriate multilevel nested design, as did other studies on cognitive measures of children's learning. Finally, none of the available studies investigated the effects of DAP beyond first grade. The effects of kindergarten DAP on social skill behaviors beyond first grade remains unexplored.

Summary of the Literature Review

Since the publication of the NAEYC's position statement (Bredekamp, 1987), developmentally appropriate practice has become so well known and been subject to a debate in the field of early childhood education. The debate is far from over since a much needed growing body of research is still being conducted. Whereas some studies reported positive effects of DAP on cognitive and psychosocial measure, others illustrated no effects or negative effects of DAP on the same measures. However, the debate cannot shadow the fact that the 1987 position statement made a significant contribution in creating an opportunity for increased conversation within and outside the field toward an improvement. NAEYC continues to answer criticism and misunderstandings or misinterpretations encountered regularly in practice by publishing new position statements and articles in various journals (e.g., Young Children). Bredekamp (1998) addresses major issues of practice that “a revised DAP and our field need to debate in the future: (1) curriculum and assessment; (2) the role of the teacher; (3) the role of culture in development; (4) attention to the individual child; and (5) relationships with families" (p.182). The need for further research, particularly on the longitudinal effects of DAP, has been addressed by many scholars. Future research regarding the psychometric limitations of previous studies would contribute to broaden the understanding of the effects of DAP on young children’s learning.

CHAPTER III

METHODOLOGY

The proposed study aims to explore the effects of developmentally appropriate classroom practices provided on children's academic development, namely reading scores, through the first four years of schooling using longitudinal data from the Early Childhood Longitudinal Study-Kindergarten Class of 1998-99. In addition, the effects of DAP across levels of gender, poverty, ethnicity, and home environment were also investigated. Considering the limitations of previous research methodology and the nature of the nested data structure, the proposed study employed an appropriate multilevel analyses to investigate the extent to which DAP is associated with children's growth in reading achievement through third grade. Furthermore, differential effects of DAP as a function of children's age, SES, ethnicity, gender, and school sector were examined. Developmentally appropriate classroom practice scores were derived from teachers' reports about several classroom activities and materials. Reading achievement was the outcome variable of interest which was measured at four time points in the study; two times in kindergarten, once in first grade, and once in third grade. The data for the multilevel analyses were drawn from the ECLS-K, a "multi-source, multi-method study that focuses on children's early school experiences beginning with kindergarten" (US Department of Education, 2000, p, 1-1). The ECLS-K provides data from a nationally representative sample of children from kindergarten through fifth grade. A total of 22,782 children throughout the US are sampled in the study and assessed children directly or indirectly in 1,277 schools which offered kindergarten programs during the 1998-99 school year. It was expected to better describe and understand the effects of kindergarten DAP with this data because when used with appropriate sample weights (provided by NCES), results from the ECLS-K data are generalizable to the United States' population of kindergarten children, teachers, and schools offering kindergarten programs in the 1998-1999 school year.

In this chapter, information is presented regarding the data source, participants, measures, and data analyses for the research questions.

Research Design

Research Questions

1. Are there differences in the acquisition of reading knowledge and skills measured by the direct and indirect (ARS) assessment batteries across kindergarten, first grade and third grade by children's exposure to more developmentally appropriate classroom practices in kindergarten?
2. Do children benefit differently from DAP in reading skills measured by the direct and indirect (ARS) assessment batteries by Age, Gender, Race, SES, and school sector?
3. Are there differences in the acquisition of reading knowledge and skills measured by the direct and indirect (ARS) assessment batteries across kindergarten, first grade and third grade by children's exposure to more developmentally appropriate classroom practices in kindergarten?
4. Does the effect of DAP on children's reading growth in kindergarten year vary by Age, Gender, Ethnicity, SES, and school sector?

Data Source

The Early Childhood Longitudinal Study-Kindergarten Class of 1998-99 (ECLS-K), conducted by National Center for Educational Statistics (NCES), is a "multi-source, multi-method study that focuses on children's early school experiences beginning with kindergarten" (US Department of Education, 2000, p, 1-1). The ECLS-K gathers data from a nationally representative sample of children from kindergarten through fifth grade. Base year sample of the ECLS-K included a total of 22,782 children who attended 1,277 schools in the school year of 1998-99, and they are assessed directly or indirectly in the fall and/or spring of kindergarten. First two waves of data were collected in the fall of 1998 and spring of 1999. There are four waves of data collection that are planned beyond kindergarten: fall and spring first grade, spring third grade and spring fifth grade. All data collection is planned to be completed in the spring of 2004 when most of the children will be in fifth grade (US Department of Education, 2000).

The ECLS-K study is an effort to better understand the entry status of kindergartners in the United States to inform educational policy and practice, regarding the fact that much of the previous literature on the status of children in schools has been focused on elementary and secondary school children (e.g., the National Assessment of Educational Progress and the National Education Longitudinal Study). Information regarding the status of the kindergarten programs in the United States and the children as they move from kindergarten through the primary grades has been very limited (West, Denton, & Germino-Hausken, 2000). Hence, results of the studies that utilize the ECLS-K database would assist to inform policies and practices regarding the diverse population of children in kindergarten. The ECLS-K provides information on the ways children are prepared for school and how schools and early childhood programs affect lives of children who attend them. The objectives and potential applications of the ECLS-K are outlined as " (1) a study of achievement in the elementary years; (2) an assessment of the developmental status of children in the United States at the start of their formal schooling and at key points during the elementary school years; (3) a cross-sectional study of the nature and quality of kindergarten programs in the United States; and (4) a study of the relationship of family, preschool, and school experiences to children's developmental status at school entry and their progress during the kindergarten and early elementary school years" (US Department of Education, 2000, p, 1-1). The ECLS-K with multivawe data enables educational policy analysts to employ techniques such as multilevel modeling to study the effects of school, classroom, and family factors on the progress of individual children.

The ECLS is a longitudinal study with two cohorts—a kindergarten cohort and a birth cohort. The birth cohort (ECLS-B) has been following a national sample of children, born in the year 2001, from birth through first grade. However, the focus of the proposed study on the effects of developmentally appropriate practices was the kindergarten cohort, ECLS-K. Being a large-scale, nationally representative, longitudinal database, ECLS-K is expected to provide insights to expand and improve early education policy and classroom practices.

There are several unique advantages of using the ECLS-K that were utilized in this study. First, ECLS-K is the first and currently the only large-scale database which is nationally representative of the status of the young children from kindergarten to the early years of elementary schooling. It provides invaluable nationally representative data on children's status at entry into school and their progress through fifth grade. Therefore, results of this study can be generalized to the U.S. population of kindergarten and first grade children, teachers, and schools.

Second, the ECLS-K provides repeated measures of the children's cognitive skills and knowledge. Data collected during the kindergarten year can serve as baseline measures to study how early experiences affect later individual development during the first three years of elementary school. The longitudinal nature of the ECLS-K enables researchers to study children's cognitive, social, and emotional growth and the trajectories of change as a function of various child, family, teacher, and school characteristics as they move through elementary school. The proposed study drew data from four waves of assessment in the fall and spring of kindergarten (fall 1998 and spring 1999), the spring of first grade (spring 2000), and the spring of third grade (spring 2002). Multiwave nature of the ECLS-K enabled to analyze the reading growth of first time kindergartners during the first four years of schooling.

Finally, the ECLS-K also provides opportunities to researchers to study the effects of a wide range of family, school, community, and individual variables on children's early success in the school system. The main interest of this study, developmentally appropriate practices, is measured by numerous classroom level variables. Such variables include, but not limited to, the frequency of various instructional approaches (e.g., whole-class teacher directed activities, child-selected activities), frequency of various classroom activities (e.g., writing from dictation, working on reading workbooks/sheets and basal reading texts), and availability of various classroom interest areas (e.g., art area, reading area with books, dramatic play area). Describing the developmentally appropriate practices using a wide range of related variables would provide a reliable and valid measure to assess the effects of classroom practices on cognitive outcomes.

Sampling Design and Sample Weights

The ECLS-K used a multistage probability sample design to reach 22,782 children in 1,277 schools who were nationally representative of the children attending kindergarten in 1998-99. First stage units were the primary sampling units (PSUs), which consisted of counties or groups of counties. Schools made the second-stage units within sampled PSUs. Finally, the third and final stage units consisted of students within those sampled schools in the previous stage.

Considering the anticipated school response rates, the average number of schools that would be selected per PSU and the target number of students to be sampled per school, a minimum number of 320 five-year-olds is determined as the criterion for a PSU to have to be eligible for selection in the study. Using the most up-to-date estimates available from the U.S. Census Bureau at the time, population estimates of five-year-olds by race-ethnicity were utilized to define a final count of 1,335 PSUs. Among those possible PSUs that met the criteria, 100 PSUs were selected for the ECLS-K.

In the next stage, public and private schools offering kindergarten programs were selected based on the data from two existing school universe files: the 1995-96 Common Core of Data2 (CCD) and the 1995-96 Private School Universe Survey3 (PSS). To be eligible for selection purposes, public schools had to have a minimum of 24 students and private schools had to have a minimum of 12 students. In all, 934 public and 346 private schools were selected for the ECLS-K. The final stage resulted in a sample of students, teachers, and parents. Although the goal of the student sample design was to obtain an approximately self-weighting sample, the effort to achieve a minimum required sample size for each targeted subpopulation required oversampling of Asian and Pacific Islanders (APIs), the only subgroup that needed to be oversampled to meet the minimum sample size goals. In the process of collecting the complete list of kindergartners enrolled, an effort was made not to exclude from the list because of disability or language problems (NCES, 2001).

Students were selected using equal probability systematic sampling from two independent sampling strata, one for APIs and one for all other non-API students. API students were sampled at 2.5 to 3.0 times the rate of non-API students in a school.

Information provided by the schools was used to locate a parent or guardian and gain parental consent for the child assessment and for the parent interview. Conducting a census of kindergarten teachers, each sampled child was linked to his or her kindergarten teacher. In cases of a child being taught more than one teacher, the “primary” teacher was assigned for the child. A more detailed and specifics discussion of the sampling design process can be found in section 5.4.2 of ECLS-K Base Year Public-Use Data File User’s Manual (NCES, 2001). This multi-stage design of sampling yielded a sample of 22,782 kindergarten children in 1,277 schools.

The first grade data collection intended to reach base year respondents who had a completed child assessment or parent interview in fall- or spring-kindergarten. However, fall-first grade data collection was limited with a 30 percent subsample. Therefore, the fall first grade data collection wave were not included in the longitudinal data analyses in this study. The spring-first grade student sample was freshened with an effort to reach current first graders who had not been enrolled in kindergarten in 1998–99. Regarding the increasing cost of additional school recruiting, a random 50 percent of children and schools were followed for fall-first grade and spring-first grade, respectively, data collection in case of children transferring to a different school than they attended in kindergarten year. Due to the 30 percent subsampling, fall-first grade sample included 5,650 children, whereas spring-first grade sample included 18,084 children, excluding freshened students (NCES, 2002).

The third grade data collection targeted base year respondents and children who were added into the sample in first grade through the freshening operation. Third grade sampling did not include a freshening process, so that no new students were added to the study. Similar to the first grade data collection, children who had transferred from their kindergarten school were followed with a slightly higher rate of random sampling than first grade. Spring-third grade sample included a total of 16,670 children, excluding the freshened students who were added into the sample in the first grade.

Sample Weights. The data collection had a stratified design structure. The ECLS-K is not a simple random sample. That is, not all schools, teachers, and children had an equal probability of selection. For instance, private schools and Asian/Pacific Islanders were oversampled to allow more powerful analyses using those subgroups. Because of this complex stratified sampling design, school-level, teacher-level, and child-level weights are provided and need to be used. Weights also allow researchers to adjust for differential nonresponses. Types of weight to be used in a specific analysis depend on three factors (NCES, 2001): (1) the type of analysis (cross-sectional or longitudinal); (2) level of analysis (child, teacher, or school); and (3) the source of the data (child assessment, parent interview, teacher questionnaire etc.). The use of these weights also enables researchers to reach estimates that are representative of the population of kindergarten children, kindergarten teachers, and schools offering kindergarten programs. Considering the longitudinal nature of the current study, following sample weights were used in data analyses: BYCW0 and C1CW0 for the kindergarten year, C3CW0 for the first grade, and C5CW0 for the third grade. These weights summed to the population of all children and produce nationally representative estimates for children who attended kindergarten in the fall of 1998. Due to the fact that the ECLS-K used a complex sample design, special procedures for estimating the statistical significance of the estimates were employed to overcome a possible inflation in the estimation of the standard errors. Using SPSS (Statistical Program for the Social Sciences), the standard errors were corrected with average root design effect (DEFT) to calculate standard errors, assuming the data were collected with a simple random sample (SRS). In the SPSS, the standard errors were corrected using DEFT. The standard error of an estimate under the actual sample design was approximated with the following formula;

$$SE_{DESIGN} = \sqrt{DEFF x Var_{SRS}} = DEFT x SE_{SRS} \qquad (1)$$

Sample

This study utilized data for the ECLS-K participants assessed in the fall of 1998, spring of 1999, spring of 2000, and spring of 2002. Two attempts were made to filter the available sample of 22,782 kindergarten children in 1,277 schools. To start with, only first time kindergartners who attended kindergarten in 1998-1999 were included in this study. In addition, only children who advanced to the first grade and the third grade in the normal pace were selected. Therefore, children who repeated either one of the kindergarten, first grade, or third grade were excluded from the study. The final sample is representative of children who attended kindergarten for the first time in the school year of 1998-1999. Next, children who had both direct and indirect measures of reading, as well as measures of DAP, at four time points were included in this study. The final sample of this study consisted of 15,498 children in kindergarten year, 11,029 children in the first grade, and 8,019 children in the third grade.

Procedures

Data collection for the first wave of the ECLS-K started in the fall of the 1998-99 school year. Computer-assisted personal interviewing (CAPI) was used to collect data from children, whereas parent data were collected by computer-assisted telephone/personal interviewing (CATI/CAPI). Data collection from teachers, school administrators, student records, and about teacher salary and benefits were conducted by using self-administered questionnaires. Teams of one field supervisor and three assessors collected data in 100 geographic work areas. Teams' responsibilities included conducting the direct child assessments and the parent interviews, distributing and collecting all school and teacher questionnaires, and completing a school facilities checklist (NCES, 2001). A total of 112 field supervisors and 343 assessors were selected from "retired teachers, former educators, people experienced in conducting assessments, or people experienced in working in schools or with school-age children" (NCES, 2001, p. 5-4) after the completion of a series of training programs.

After seeking and receiving endorsements of many national associations and organizations representing parents, school administrators, teachers, and schools, the chief state school officer of each of the 41 states that contained ECLS-K sampled schools were contacted to explain the objectives of the study and the data collection procedures in January 1998. Approvals from the state level were followed by contacting the district superintendents the Catholic dioceses and Archdioceses in the sample to obtain permission to contact schools. School administrators were contacted after getting the permissions from the districts and dioceses beginning in March 1998. Once the commitments from the school administrator were received, two appointments for two visits were scheduled; one for the pre-assessment to select the sample of children and one for the child assessment. In terms of eligible children, the weighted response rate for public schools was 70.1 percent, whereas the response rate for Catholic schools was 83.0 and for other private schools was 56.2, indicating a total of 65.9 percent response rate for all private schools (NCES, 2001).

Fall kindergarten data collection was conducted in 1998 from September to December. Although it seems like a large span of time, previous reports from the ECLS-K ensure that 80% of the assessment were conducted between early October and mid-November (Rathbun, West, & Germino-Hausken, 2004). The spring kindergarten data were collected between March 1999 and June 1999. Other two waves of data collection were conducted in spring first grade in 2000 from March to July and in spring third grade in 2002 form March to July, with 80% of the assessment at each wave conducted between early April and late May (NCES, 2001; 2002; 2004; Rathbun, West, & Germino-Hausken, 2004).

Conducting the Direct Child Assessment. Prior to the child data collection, children whose home language was other than English (as determined by school records) were identified and administered the Oral Language Development Scale (OLDS), a measure of basic English proficiency, to ascertain language minority children could be validly and reliably assessed. The OLDS is a subset of the PreLAS 2000 (Duncan & DeAvila, 1998) that measures receptive and expressive language in English. Only 15 % of sampled children were identified for the OLDS screening test (NCES, 2001). Of these children, 42 % whose home language was Spanish, and 61 % whose home language was other than English or Spanish were at or above the cut score, therefore qualified to take reading battery in English. Children who performed below the cut score, were administered the Spanish version of the OLDS and some other direct cognitive measures, if their primary home language was Spanish. However, the major focus of this study, children's early reading and literacy development, was assessed only by tests administered in English. This process eliminated 1,567 children in the fall of kindergarten and 350 in the spring of first grade from being administered the reading battery in English.

Direct child assessment was conducted during a 14-week field period. Following the procedures for meeting children that were agreed upon during the preassessment visit at the school, assessment teams prepared the room, which was normally a school classroom or library, and administered the tests in a 50- to 70-minute session per child. Upon the completion of the assessment, children were signed back into the classroom by the field staff. An effort was made "to conduct the direct child assessments at about the same point in time from the beginning of school year and at the end of the year to increase the chances that children's exposure to instruction was about the same for all children" (NCES, 2001, p. 5-12).

Conducting the Parent Interview. Parent interviews were conducted primarily by telephone, using a CATI by field staff during a 35- (fall first grade) to 62- (spring third grade) minute session. In cases of parents not having a telephone, the interview was conducted in person. This procedure yielded 1 to 3 % of in person interview in various round of data collection. The interviews were conducted primarily in English. Parents who spoke only Spanish, Lakota, Hmong, or Chinese were interviewed by a bilingual interviewer in the language they spoke. Approximately seven percent of the parents were interviewed in a language other than English; of these interviews, 94 percent were conducted in Spanish. Due to language problems, one percent of parents were excluded from being interviewed.

Teacher Data Collection. During the preassessment visit, teachers in the sampled schools received the self-administered teacher questionnaires A, B, and C and were asked to return them to the field supervisors at the next assessment visit. Telephone contacts and additional visits to the school by the field supervisor were made to collect completed teacher questionnaires that were not available during the assessment visits. For each completed Teacher Questionnaire C, which was the only child-specific questionnaire, teachers were reimbursed five dollars in kindergarten year and seven dollars in the next rounds of data collection.

Measures

Cognitive Outcome Measures. The Early Childhood Longitudinal Study-Kindergarten Class of 1998-99 provides various types of scores on various types of scales to describe children's cognitive and social development during their kindergarten year. These scores are for the direct cognitive assessment, the academic rating scale (ARS), the psychomotor assessment, and the social rating scale (SRS). As a part of the direct cognitive assessment, reading, mathematics, and general knowledge skills of young children are assessed in the fall and spring of kindergarten, spring of first grade, and spring of third grade. Because the proposed study is focused on the factors, particularly on developmentally appropriate practices, associated with reading achievement of young children from kindergarten through third grade, one direct and one indirect measure of children's reading and literacy were used as the measures of the outcome. Thus, reading achievement is the outcome variable of interest which is measured at four time points in the study. Below, descriptions of the scores, variable names, and variable descriptions are provided.

Direct Cognitive Assessment. There are three areas of direct cognitive assessment in the ECLS-K; reading, mathematics, and general knowledge. Whereas the same reading and mathematics battery was used in kindergarten and first grade, a new battery was developed with additional items that were designed to assess more advanced skills, regarding the expected academic development of children. In the first stage, children received a 12- to 20-item routing test. Results of this routing test guided the selection of one of several alternative second-stage forms that consisted of items with an appropriate level of difficulty tailored to the child's current level of ability.

Direct Cognitive Assessment of Reading. Before starting to collect data with direct cognitive measure, a group of experts that included experts in child development, primary education, and testing methodology, along with their counterparts at NCES, reviewed commercially available tests in the field and indicated that there were no "off-the-shelf" tests that met the requirements. This yielded a recommendation from the group that an individually administered and adaptive test to be developed or adapted. It is proposed that "since young children are not experienced test-takers, individual administration could provide more sensitivity to each child's needs than a group-administered test" (Rock & Pollack, 2002, 2-2). Furthermore, it was also recommended that the tests be adaptive to be able to test each child more appropriately using items that were tailored to his or her level of achievement, regarding some psychometric test developing criteria (e.g., floor and ceiling effects).

The framework of the ECLS-K reading specifications was drawn mainly from the Reading Framework for the 1992 and 1994 NAEP. However, because the earliest grade targeted in the NAEP was fourth grade, the ECLS-K assembled a group of experts consisted of literacy curriculum specialists, test item writers from the Educational Testing Service (ETS), and kindergarten through second grade teachers to review the proposed framework and item pool, in addition to developing new assessment items and adapting existing items from published tests. In an effort to develop a new reading assessment test, NCES sought permission to barrow or adapt items from published tests including the Peabody Individual Achievement Test-Revised (PIAT-R), Peabody Picture Vocabulary Test-Revised (PPVT-R), the Primary Test of Cognitive Skills (PTCS), the Test of Early Reading Ability (TERA-2), the Test of Early Mathematics Ability (TEMA-2), and the Woodcock- Johnson Tests of Achievement-Revised (WJ-R) (Rathbun, West, & Germino-Hausken, 2004; Rock & Pollack, 2002).

The kindergarten and first-grade reading assessments in the ECLS-K covered a range of items measuring basic skills (print familiarity, letter recognition, beginning and ending sounds, rhyming sounds, "sight" word recognition), vocabulary (receptive vocabulary), and comprehension (listening comprehension, words in context). The third grade reading assessment was, instead, designed to measure skills in phonemic awareness, single word decoding, vocabulary (reading), and passage comprehension. Despite the different grade levels, comprehension skills for the K-1 and third grade reading assessments were related in some level and all mainly reflect initial understanding, developing interpretation, personal reflection, and demonstrating critical stance. These reading assessments contained varied proficiency levels, characterizing children's progression of skills and knowledge in language and literacy. Specifically, if a child correctly responds to all items in a proficiency level, it is considered that he or she mastered the skills in that level. Furthermore, if a child masters one of the advanced proficiency levels, he or she was also thought to pass the items related to the earlier proficiency levels.

The ECLS-K reading assessment for the kindergarten and first grade levels included five proficiency levels as follows: (1) identifying upper- and lower-case letters of the alphabet by name; (2) associating letters with sounds at the beginning of words; (3) associating letters with sounds at the end of words; (4) recognizing common "sight" words; and (5) reading words in context. The third grade proficiency levels also include five proficiency levels as (1) recognizing common "sight" words; (2) reading words in context; (3) making inferences using cues that were directly stated with key words in text (literal inference); (4) identifying clues used to make inferences (extrapolation), and using personal background knowledge combined with cues in a sentence to understand use of homonyms; and (5) demonstrating understanding of author's craft and making connections between a problem in the narrative and similar life problems (evaluation). As is seen, the K-1 and the third grade assessments were partially overlapped in two proficiency levels. The last two (advanced) proficiency levels of recognizing common "sight" words and reading words in context in kindergarten and first grade were retained in the third grade assessment "in order to link the kindergarten and first-grade assessment and the third-grade assessment for scaling purposes" (Rathbun, West, & Germino-Hausken, 2004, p. 4). Besides that, three higher proficiency levels were added at the third grade level as literal inference, extrapolation, and evaluation.

There are five different types of scores that were computed in the ECLS-K to portray children's achievement on the direct cognitive assessment (NCES, 2001, p. 3-1): (1) number right scores; (2) item response theory (IRT) scores; (3) standardized scores; (4) criterion-referenced proficiency level and (5) proficiency probability scores. For the purposes of this longitudinal study, a composite variable on an IRT based scale in reading, that combines all those skills mentioned above, was utilized in a continuous form.

Reading scores based on the full set of test items were calculated using IRT procedures. IRT uses the pattern of right, wrong, and omitted responses to the items actually administered in a test and the difficulty, discriminating ability, and "guess-ability" of each item to place each child on a continuous ability scale. The items in the routing test, plus a core set of items shared among the different second stage forms, made it possible to establish a common scale. It is then possible to estimate the score the child would have achieved if all of the items in all of the test forms had been administered. IRT has several advantages over raw number-right scoring. By using the overall pattern of right and wrong responses to estimate ability, IRT can compensate for the possibility of a low ability student guessing several hard items correctly. Omitted items are also less likely to cause distortion of scores, as long as enough items have been answered right and wrong to establish a consistent pattern. Unlike raw scoring, which, in effect, treats omitted items as if they had been answered incorrectly; IRT procedures use the pattern of responses to estimate the probability of correct responses for all test questions. Finally, IRT scoring makes possible longitudinal measurement of gain in achievement over time, even though the tests administered are not identical at each point (US Department of Education, 2000).

Reliability statistics for instruments in each subject area for each round of data collection were provided and NCES (2004, p.3-23) states that "for the IRT-based scores, the reliability of the overall ability estimate, theta, is based on the variance of repeated estimates of theta". Reliability estimates for the reading IRT scores were .93, .95, .96, and .94, for the fall kindergarten, spring kindergarten, spring first-grade, and spring third-grade, respectively.

Indirect Cognitive Assessment of Reading by the ARS. In addition to direct cognitive assessment of reading, the ECLS-K also used an indirect nonstandardized cognitive assessment called the Academic Rating Scale. The main purpose of the ARS measure was to overlap and enhance the information gathered through the direct cognitive assessment. The ARS was designed to assess teachers' evaluations of student academic performance on the same broad curricular domains, reading, general knowledge (science and social studies), and mathematics as in the direct assessments. Although both direct and indirect assessments were designed to collect information about children's academic skills and knowledge on the exact same subject areas, there were some variations between these two measures.

Using standardized tests to measure the effects of developmentally appropriate practices has long been criticized because the nature of DAP is thought to be contrary to the philosophy of standardized tests (e.g., Bredekamp, 1997; Hart, Burts, & Charlesworth, 1997; Kostelnik, Soderman, & Whiren, 2004; Miller, 1996). Some insignificant or no effects found in the studies were also explained by this contradiction, based on the fact that whereas DAP is more process oriented, standardized tests generally measure the final product. Measuring the effects of DAP with standardized tests considered to be inappropriate by some researchers in the field. Therefore, the ARS in this study was utilized to alternatively assess the effects of DAP, along with various child and classroom characteristics, on young children's reading achievement during the first four years of schooling. The ARS mostly focused on both the process and products of children's learning, whereas, due to time and space limitations, the direct cognitive assessments measured only the products of children's learning. Constraints of standardized testing in direct cognitive measures did not limit the development of the ARS items. The ARS was designed to include skills, knowledge, and behaviors that reflect a broader sampling of the most recent national curriculum standards and guidelines from early childhood professionals and researchers.

Similar to the direct cognitive measures, the level of skills measured by the ARS also increases by grade. Teachers were provided with examples to distinguish between the levels of difficulty of an item on the ARS. Third grade items were designed to represent a higher level of difficulty than the items that belong to the previous grades. For example, measuring the reading fluency, whereas the first grade item asked whether the child read first grade book fluently with words in meaningful phrases, the third grade equivalent item asked whether the child read fluently including words with three or more syllables (NCES, 2004).

In all subject areas, as well as the reading and literacy section, children were rated by their teachers on a five-point rating scale from “Not Yet”, “Beginning”, “In Progress”, “Intermediate”, to “Proficient”. A code of Not Applicable (N/A) was given if the child was not presented with a certain level of skill, knowledge, or behavior in the classroom yet. Correspondingly, scores on the ARS Language and Literacy ranged between a low of one and a high of five.

Different sets of item ratings were developed for the fall-kindergarten, spring-kindergarten, spring first-grade, and spring third-grade ARS instruments. Although the item stems are similar across grades, the level of expected performance was increased between grades. There was sufficient overlap of identical items in the fall- and spring-kindergarten forms that a common calibration could be carried out, therefore, it was possible to run a longitudinal or gain score analysis in kindergarten. However, the first-grade and third-grade items differed from the kindergarten items and were calibrated on a on a different metric. Consequently, no analysis could have been conducted using ARS gain scores across grades.

Similar to the direct cognitive measures, IRT-based ARS scores were computed, using the Rasch Rating Scale Model, to compare performance of students and to compensate for nonresponses on some items. Reliability estimates for the IRT-based ARS Language and Literacy scores were .87, .91, .94, and .95, for the fall kindergarten, spring kindergarten, spring first-grade, and spring third-grade, respectively.

Measures of Student Characteristics

Age. This variable is provided by the ECLS-K data sets and was computed by determining the number of days between the date when direct child assessment was completed (e.g., the spring-first grade source variables were C4ASMTMM, C4ASMTDD, C4ASMTYY) and the child’s date of birth (DOBMO, DOBDA, DOBYR). Children’s age in months was computed in kindergarten and first grade by dividing the total number of days by 30. However, the age variable in third grade was categorical with five levels. Therefore, the age variable at this grade was dummy coded as YOUNGER and OLDER. The reference group was the 9 years old, which was the appropriate chronological age for third grade. In this study, variable of R1_KAGE, R2_KAGE, R4_AGE, and R5AGE were used as indicators of children's age at the time of assessment.

Gender. This variable was derived from the parent interview. If it was missing or the reported gender was different in the fall-kindergarten parent interview and spring-kindergarten parent interview, data from the field management system was used.

Race/Ethnicity. The race/ethnicity is a composite variable that was derived from parent interviews as two separate variables: race and ethnicity. Parents were able to selects more than one race. First, parents indicated the child's race using five race categories: White, Black or African American, American Indian or Alaskan Native, Asian, Native Hawaiian or other Pacific Islander. Additionally, one more dichotomous variable was created to indicate if the child was multiracial without specifying the race. Secondly, parents were asked to indicate the child's ethnicity using the following categories: White, non Hispanic; black or African American, non-Hispanic; Hispanic, race specified; Hispanic, no race specified; Asian; Native Hawaiian or other Pacific Islander; American Indian or Alaskan Native. In this study, a composite of following race/ethnicity categories were used: (1) White, (2) Black, (3) Hispanic, (4) Asian, and (5) Other (which included Pacific Islanders, American Indians, Alaska Natives, and non-Hispanic multiracial children). The Caucasion (White) group was selected as the reference group and coded as zero (0).

Socioeconomic Status (SES). In this investigation a continuous measure of family socioeconomic status (WKSESL) was used as a child level variable. The SES variable was derived from two parent interviews in kindergarten year. The SES was computed using the following components: Father/male guardian's education; mother/female guardian's education; father/male guardian's occupation; mother/female guardian's occupation; and household income. The parent's occupation reflected the average of the 1980 and 1989 General Social Survey prestige scores of the occupation (NCES, 2001). Because there were 2% (mother's education) to 28 % (household income) missing values for each component, a hot deck imputation methodology was used to impute for missing values. That is, the value reported by a respondent for a particular item is given to a similar person who failed to respond to that question. Then, each component converted to a *z*-score with a mean of 0 and a standard deviation of one. The average of all five components created the continuous WKSESL variable that ranged from –4.75 to 2.75.

Measures of Classroom Environment

Developmentally Appropriate Practices. The DAP variable was derived from several teacher reported indicators that describe teachers' classroom practices and the materials available in their perspective classrooms. This was a continuous variable reflecting the level of DAP in a classroom. Higher scores on this variable reflected that the child was in a more child-centered classroom.

The first group of variables used to describe DAP consisted of two activities teachers reported: teacher-directed whole class activities and child-selected activities. There were two other activities presented under this category: teacher-directed small group activities and teacher-directed individual activities. However, because these two latter items can be categorized under either DAP or DIP depending on the individual differences in application, they were not included for this analysis. Teachers indicated how much time they spent on these activities during a typical day. Responses for these items on a five-point scale ranged from (1) no time, (2) half hour or less, (3) about one hour, (4) about two hours, to (5) three hours or more. A preliminary investigation of the descriptive statistics of these variables indicated that the total hours reported on all activities by teachers differed, even after controlling for the school day (half-day or full-day). In an effort to eliminate possible reporting bias and school day differences, total time for all four activities was computed and then the proportion of the time spent on each activity was calculated. Therefore, for the two selected items, children's time spent on each activity was indicated in percentages.

The second group of variables consisted of the areas or centers for activities in a classroom. Teachers reported whether their classrooms had any of the 11 interest areas or centers. These areas or centers were reading area with books, listening center, writing center or area, pocket chart or flannel board, math area with manipulatives, area for playing with puzzles and blocks, water or sand table, computer area, science or nature area with manipulatives, dramatic play area or corner, and art area.

The third group of variables consisted of children's use of four materials in the classroom: art materials, musical instruments, costumes for creative dramatics/theater, and cooking or food related items. Teachers reported how frequently children use these materials in their classroom on a six-point scale that ranged from (1) never, (2) once a month or less, (3) two or three times a month, (4) once or twice a week, (5) three or four times a week, to (6) daily.

The continuous DAP variable was created by a confirmatory factor analysis (CFA). Two external criteria were also used to verify the classification and the predictive validity of the measure. Class evaluation practices of teachers', which were theoretically related to DAP, were used to validate the clustering process. These evaluation practices are as follows: (1) Individual child's achievement relative to local, state, or professional standards; and (2) individual improvement or progress over past performance. As it was hypothesized, the correlations between DAP and two evaluation practices were significant as $r = -.19$ and $r = .21$, respectively, in kindergarten. The same correlations were estimates as $r =-.22$ and $r = .24$ for first grade and $r = -.18$ and $r = .22$ for the third grade. All the correlations reported were statistically significant at $p< .05$ level.

Confirmatory factor analysis for the DAP model

Teachers' developmentally appropriate practice as a latent variable was derived utilizing a confirmatory factor analyses (CFA). Based on grade levels, three DAP models were designed with three latent constructs. The kindergarten DAP model includes 16 observed variables whereas the first and third grade DAP models contains 14 observed items (see Figures 1 & 2 in Appendix). As can be seen from the models, each one is a second-order model, represented by observed and latent constructs. At the first and third grade levels, two items were excluded from models because of their factor loadings did not meet the criteria, which is .40 or more.

In assessing models, the chi-square statistic and three goodness-of-fit statistics were considered: Comparative fit index (CFI), Tucker-Lewis index (TLI), and Root mean square error of approximation (RMSEA). Hu and Bentler (1999) advocated values greater than 0.95 for CFI and TLI as a minimum value that can be used to conclude that there is a relatively good fit between the hypothesized model and the data. RMSEA values of 0.06 or less indicate adequate fit. The results of the all grade level models revealed that the hypothesized second-order latent factor models moderately fit the data (see Table 48 in Appendix). Even though models did not present good fit in terms of chi-square statistics ($p>.05$), CFI and TLI values were very close to cutoff value of .95. Moreover, all RMSEA values were estimated as lower than the criterion of .06.

Given the promising fit indices, satisfactory factor loadings, and the theoretical support for the content, the factor scores on the latent DAP factor that were generated from the confirmatory factor model were used as a classroom level predictor in further HLM analyses.

Statistical Analyses

Being a large-scale, nationally representative, longitudinal database, ECLS-K is expected to provide insights to expand and improve early education policy and classroom practices. This study was an attempt to explore the effects of developmentally appropriate classroom practices provided in kindergarten and earl years of elementary school on children's academic achievement, namely reading scores. Considering the limitations of previous research methodology (Van Horn & Ramey, 2003), such as the violations of the assumptions of independence of observations that is essential to regression and ANOVA models (Tabachnick & Fidel, 1996), the proposed study employed an appropriate multilevel analyses, that considers the nested data structure, to investigate the extent to which DAP is associated with children's development in reading achievement from kindergarten through 3rd grade. Furthermore, differential effects of DAP as a function of children's age, SES, race, gender, and school sector were examined.

Reading achievement is the outcome variable of interest which was measured at four time points in the study; two times in kindergarten, once in first grade, and once in third grade. Hierarchical Linear Modeling (HLM), a multilevel analysis strategy, was utilized in the study. HLM allows for the analysis of hierarchically structured data, that is where individuals are nested within groups and groups are nested within larger groups.

The guidelines of Raudenbush and Bryk (2002) were utilized in all HLM analyses using the HLM 5.04 version. Analyses started with the estimation of an unconditional model. Then, the model was expanded by the inclusion of individual-level (student) and classroom level predictors. All variables that do not have a meaningful zero (0) were grand mean centered at all levels.

Cross-Sectional Multilevel Analysis. Six cross-sectional multilevel models were constructed to address the first and second research questions. Each model specified by grade level for both outcome variables was analyzed based on a three-level hierarchical linear approach.

Research Questions:

1. Are there differences in the acquisition of reading knowledge and skills measured by the direct and indirect (ARS) assessment batteries across kindergarten, first grade and third grade by children's exposure to more developmentally appropriate classroom practices in kindergarten?
2. Do children benefit differently from the DAP in reading skills measured by the direct and indirect (ARS) assessment batteries by Age, Gender, Race, SES, and school sector?

The current model followed the three-stage approach of multi-level analysis. In the first stage, the analysis produced the null model with no independent variables at the student and the classroom levels. With only the student level outcome measure, herein reading IRT scores, this model is similar to random effect ANOVA model, providing a measure of the variances within and between classrooms for reading achievement. At the second stage (random coefficients model), student-level variables were added to null model to determine whether their relationship with achievement varied significantly across classrooms. The last stage refers to full model (intercepts and slopes as outcomes). Classroom and school level variables were added to the model. As mentioned earlier, separate models were used for each grade level, representing kindergarten, first, and third grades of data. A detailed explanation of the HLM modeling is provided in the Results chapter.

Longitudinal Data Analysis. Following a similar approach, a three-level growth modeling was used to investigate research questions 3 and 4 using the Kindergarten level dataset. A detailed explanation of the HLM growth modeling is also provided in the Results chapter.

Research questions:

3. Is there evidence for systematic change and individual variability in reading growth assessed through the direct and indirect (ARS) assessment batteries associated with teachers' use of developmentally appropriate practices over kindergarten year?
4. Does the effect of DAP on children's reading growth in kindergarten year vary by Age, Gender, Race, SES, and school sector?

CHAPTER IV

RESULTS

This chapter displays the results of data analyses that explored the effects of developmentally appropriate classroom practices on children's academic growth, namely reading scores, through the first four years of schooling using longitudinal data from the Early Childhood Longitudinal Study-Kindergarten Class of 1998-99. Outcome variables of interest were reading scores measured in two different ways; a direct standardized measure of reading IRT scores and an indirect teacher reported reading score assessed by the Academic Rating Scale. After providing the descriptive statistics at each grade, a series of Hierarchical Linear Models (HLM) that were developed to evaluate the effects of DAP at each grade level, along with two longitudinal analyses in the kindergarten year, are demonstrated. Results of the HLM analyses were organized in regard to the grade level, starting with cross-sectional kindergarten analyses, followed by cross-sectional first grade analyses, cross-sectional third grade analyses, and finally longitudinal kindergarten analyses. All three-level HLM analyses begin with the Random Effects Analysis of Variance (ANOVA) model to examine the existence of differences in the mean reading performance among classrooms and the amount of variation that is within and between classrooms. Next, a Random-Coefficient model is developed to examine the variability of intercepts and slopes across classrooms. Lastly, the Intercepts- and Slopes-as-Outcomes model is demonstrated to build an explanatory model to account for the variability of the intercepts and slopes. Same procedures are followed for the three-level longitudinal analyses of the effects of DAP during the kindergarten year. The statistical analyses were conducted using software programs of SPSS 10.0, HLM 5.04, and Mplus.

HLM Application

Utilizing various HLM applications, this study investigated the relationships between seven student-level independent variables (i.e., age, SES, gender, and four dummy codes of race/ethnicity), one classroom-level independent variable (i.e., developmentally appropriate practice), one school level variable (i.e., school sector coded as public and private), and two student-level outcome variables (i.e., reading IRT score and reading ARS score). In the first step of the cross-sectional analyses, the Random Effects ANOVA model was tested to provide an estimate of the grand mean (average reading scores across all classrooms) and to partition the total variation in performance into between and within classrooms. The Random-Coefficient model was run to gain information about the variability of intercepts and slopes across classrooms in the second phase. As the last step, the Intercepts- and Slopes-as-Outcomes model presented an exploratory model to account for the variability of the intercepts and slopes.

Descriptive Statistics

Descriptive analyses of the data were conducted to describe the reading performance of first-time kindergartners at four waves of data collection. The average reading IRT score in spring kindergarten was 22.62, followed by scores of 32.70 in spring kindergarten, 57.47 in spring first grade, and 112.04 in spring third grade. The average reading ARS score in spring kindergarten was 2.62, followed by scores of 3.40 in spring kindergarten, 3.49 in spring first grade, and 3.43 in spring third grade. Scores on reading IRT and reading ARS scales by gender are presented in Table 2 and Table 3, respectively. Girls consistently outperformed their boy counterparts at all four time points on both reading measures. They started kindergarten with higher scores on both reading measures and retained that advantage through third grade. However, these differences were not large in terms of effects sizes, as all differences were smaller than .15 standard deviation units.

Table 2
Reading IRT Scores by Gender

	Fall Kindergarten		Spring Kindergarten		Spring First Grade		Spring Third Grade	
	M	SD	M	SD	M	SD	M	SD
Male	21.50	10.24	31.80	10.42	56.41	13.07	110.75	18.16
Female	23.26	8.47	33.61	10.25	58.52	12.51	113.25	17.32
Total	22.62	8.56	32.70	10.37	57.47	12.83	112.04	17.77

Note. M= Mean; SD= Standard deviation.

Table 3
Reading ARS Scores by Gender

	Fall Kindergarten		Spring Kindergarten		Spring First Grade		Spring Third Grade	
	M	SD	M	SD	M	SD	M	SD
Male	2.54	.78	3.31	.79	3.38	.89	3.32	.83
Female	2.69	.77	3.49	.78	3.59	.90	3.53	.83
Total	2.62	.78	3.40	.79	3.49	.90	3.43	.84

Note. M= Mean; SD= Standard deviation.

Pearson product moment correlations between the reading achievement measure and the DAP variables at four time points are presented in Table 4.

.

Table 4

Correlations Between Outcome Measures and DAP Scores

	c1rscale	c2rscale	t1arslit	t2arslit	c4rrscal	t4arslit	c5r2rscl	t5arslit	dap_k	dap_1
c1rscale										
c2rscale	.811*									
t1arslit	.192*	.172*								
t2arslit	.312*	.459*	.173*							
c4rrscal	-.012	-.020*	.007	-.006						
t4arslit	-.014	-.020*	.009	-.016	.559*					
c5r2rscl	-.033*	-.036*	-.006	.010	-.004	.007				
t5arslit	-.011	-.006	-.011	-.014	.008	.008	.262*			
dap_k	.069*	.055*	.059*	.037*	.019*	.017	.027*	.025*		
dap_1	-.016	-.007	.003	.013	-.026*	.009	-.011	-.005	-024*	
dap_3	-.013	-.001	-.022*	.011	-.026*	.001	-.003	.033*	.021*	-.001

* Correlation is significant at the 0.05 level (2-tailed).

Note: c1rscale=IRT score at time 1; c2rscale= IRT score at time 2; c4rrscal= IRT score at time 4; c5r2rscl= IRT score at time 5 t1arslit= ARS score at time 1; t2arslit= ARS score at time 2; t4arslit= ARS score at time 4; t5arslit ARS score at time 5; dap_K: Kindergarten DAP score; dap_1= First grade DAP score; dap_3= Third grade ARS score.

Descriptive statistics of HLM analyses sorted by the level of analysis (child, classroom, and school) for each independent variable at each data collection point can be found in Tables 5 through 12.

Table 5

Descriptive Statistics for Reading IRT Score Analyses in Spring Kindergarten

Level	Variables	Type	N	M	SD
	Outcome				
	READ_IRT	Continuous	15000	32.61	10.37
	Predictors				
Level 1 (Student Level)	MALE	Dummy Coded (Male=1, Female=0)	15000	.50	.50
	AGE	Continuous (in months)	15000	68.25	4.12
	SES	Continuous	15000	.08	.78
	BLACK	Dummy coded (Black=1 vs Others=0)	15000	.15	.35
	HISPANIC	Dummy coded (Hispanic=1 vs Others =0)	15000	.14	.35
	ASIAN	Dummy coded (Asian=1 vs Others =0)	15000	.05	.22
	OTHER	Dummy coded (Other=1 vs Others =0)	15000	.06	.23
Level 2 (Classroom Level)	DAP	Continuous	2893	-.09	2.52
Level 3 (Classroom Level)	PUBLIC	Dummy Coded (Public=1 vs. Private=0	941	.78	.42

Note. N= Sample size; M= Mean; SD= Standard deviation.

Table 6

Descriptive Statistics for Reading ARS Score Analyses in Spring Kindergarten

Level	Variables	Type	N	M	SD
	Outcome				
	READ_ARS	Continuous	15498	3.39	.79
	Predictors				
Level 1 (Student Level)	MALE	Dummy Coded (Male=1 vs. Female=0)	15498	.50	.50
	AGE	Continuous (in months)	15498	74.44	4.19
	SES	Continuous	15498	.04	.79
	BLACK	Dummy coded (Black=1 vs. Others =0)	15498	.14	.35
	HISPANIC	Dummy coded (Hispanic=1 vs. Others =0)	15498	.17	.37
	ASIAN	Dummy coded (Asian=1 vs. Others =0)	15498	.07	.25
	OTHER	Dummy coded (Other=1 vs. Others =0)	15498	.04	.20
Level 2 (Classroom Level)	DAP	Continuous	3042	-.15	2.52
Level 3 (Classroom Level)	PUBLIC	Dummy Coded (Public=1 vs. Private=0	972	.77	.42

Note. N= Sample size; M= Mean; SD= Standard deviation.

Table 7

Descriptive Statistics for Longitudinal Reading IRT Score Analyses in Kindergarten

Level	Variables	Type	N	M	SD
	Outcome				
Level 1 Time	READ_IRT	Continuous	29822	27.66	10.76
	YEAR	Continuous (Age at the time of testing in months)	14882	99.99	159.90
	Predictors				
Level 2 (Student Level)	MALE	Dummy Coded (Male=1 vs. Female=0)	14882	.50	.50
	AGE	Continuous (in months)	14882	68.28	4.12
	SES	Continuous	14882	.09	.78
	BLACK	Dummy coded (Black=1 vs. Others=0)	14882	.15	.36
	HISPANIC	Dummy coded (Hispanic=1 vs. Others=0)	14882	.13	.33
	ASIAN	Dummy coded (Asian=1 vs. Others=0)	14882	.06	.23
	OTHER	Dummy coded (Other=1 vs. Others=0)	14882	.05	.21
Level 3 (Classroom Level)	DAP	Continuous	968	-.35	3.18

Note. N= Sample size; M= Mean; SD= Standard deviation.

Table 8

Descriptive Statistics for Longitudinal Reading ARS Score Analyses in Kindergarten

Level	Variables	Type	N	M	SD
	Outcome				
Level 1 Time	READ_IRT	Continuous	28507	3.01	.88
	YEAR	Continuous (Age at the time of testing in months)	28507	99.82	159.80
	Predictors				
Level 2 (Student Level)	MALE	Dummy Coded (Male=1 vs. Female=0)	14231	.50	.50
	AGE	Continuous (in months)	14231	68.22	4.11
	SES	Continuous	14231	.04	.80
	BLACK	Dummy coded (Black=1 vs Others =0)	14231	.14	.35
	HISPANIC	Dummy coded (Hispanic=1 vs Others =0)	14231	.17	.37
	ASIAN	Dummy coded (Asian=1 vs Others =0)	14231	.07	.25
	OTHER	Dummy coded (Other=1 vs Others =0)	14231	.04	.20
Level 3 (Classroom Level)	DAP	Continuous	961	-.32	3.17

Note. N= Sample size; M= Mean; SD= Standard deviation.

Table 9

Descriptive Statistics for Reading IRT Score Analyses in Spring First Grade

Level	Variables	Type	N	M	SD
	Outcome				
	READ_IRT	Continuous	10934	57.47	12.84
	Predictors				
Level 1 (Student Level)	MALE	Dummy Coded (Male=1 vs. Female=0)	10934	.50	.50
	AGE	Continuous (in months)	10934	86.76	4.05
	SES	Continuous	10934	.06	.79
	BLACK	Dummy coded (Black=1 vs. Others =0)	10934	.13	.33
	HISPANIC	Dummy coded (Hispanic=1 vs. Others =0)	10934	.15	.35
	ASIAN	Dummy coded (Asian=1 vs. Others =0)	10934	.06	.24
	OTHER	Dummy coded (Other=1 vs. Others =0)	10934	.04	.20
Level 2 (Classroom Level)	DAP	Continuous	3256	.14	1.98
Level 3 (Classroom Level)	PUBLIC	Dummy Coded (Public=1 vs. Private=0	1215	.81	.40

Note. N= Sample size; M= Mean; SD= Standard deviation.

Table 10

Descriptive Statistics for Reading ARS Score Analyses in Spring First Grade

Level	Variables	Type	N	M	SD
	Outcome				
	READ_ARS	Continuous	11029	3.49	.90
	Predictors				
Level 1 (Student Level)	MALE	Dummy Coded (Male=1 vs. Female=0)	11029	.50	.50
	AGE	Continuous (in months)	11029	86.74	4.05
	SES	Continuous	11029	.04	.80
	BLACK	Dummy coded (Black=1 vs Others =0)	11029	.12	.33
	HISPANIC	Dummy coded (Hispanic=1 vs Others =0)	11029	.16	.37
	ASIAN	Dummy coded (Asian=1 vs Others =0)	11029	.06	.24
	OTHER	Dummy coded (Other=1 vs Others =0)	11029	.04	.20
Level 2 (Classroom Level)	DAP	Continuous	3284	.14	1.98
Level 3 (Classroom Level)	PUBLIC	Dummy Coded (Public=1 vs. Private=0	1209	.81	.40

Note. N= Sample size; M= Mean; SD= Standard deviation.

Table 11
Descriptive Statistics for Reading IRT Score Analyses in Third Grade

Level	Variables	Type	N	M	SD
	Outcome				
	READ_IRT	Continuous	8019	112.04	17.77
	Predictors				
Level 1 (Student Level)	MALE	Dummy Coded (Male=1 vs. Female=0)	8019	.48	.50
	YOUNGER	Dummy Coded (Younger than 9=1 vs. 9 year old=0)	8019 8019	.26	.44
	OLDER	Dummy Coded (Older than 9=1 vs. 9 year old=0)	8019 8019	.25	.43
	SES	Continuous	8019	.10	.78
	BLACK	Dummy coded (Black=1 vs. Others =0)	8019	.09	.29
	HISPANIC	Dummy coded (Hispanic=1 vs. Others =0)	8019	.14	.35
	ASIAN	Dummy coded (Asian=1 vs. Others =0)	8019	.06	.24
	OTHER	Dummy coded (Other=1 vs. Others =0)	8019	.04	.19
Level 2 (Classroom Level)	DAP	Continuous	2974	.08	1.42
Level 3 (Classroom Level)	PUBLIC	Dummy Coded (Public=1 vs. Private=0	1490	.82	.38

Note. N= Sample size; M= Mean; SD= Standard deviation.

Table 12

Descriptive Statistics for Reading ARS Score Analyses in Third Grade

Level	Variables	Type	N	M	SD
	Outcome				
	READ_ARS	Continuous	7728	3.43	.84
	Predictors				
Level 1 (Student Level)	MALE	Dummy Coded (Male=1 vs. Female=0)	7728	.48	.50
	YOUNGER	Dummy Coded (Younger than 9=1 vs. 9 year old=0)	7728 7728	.26	.44
	OLDER	Dummy Coded (Older than 9=1 vs. 9 year old=0)	7728 7728	.26	.44
	SES	Continuous	7728	.10	.78
	BLACK	Dummy coded (Black=1 vs. Others =0)	7728	.09	.29
	HISPANIC	Dummy coded (Hispanic=1 vs. Others =0)	7728	.14	.35
	ASIAN	Dummy coded (Asian=1 vs. Others =0)	7728	.06	.24
	OTHER	Dummy coded (Other=1 vs. Others =0)	7728	.04	.20
Level 2 (Classroom Level)	DAP	Continuous	2899	.08	1.42
Level 3 (Classroom Level)	PUBLIC	Dummy Coded (Public=1 vs. Private=0	1464	.82	.38

Note. N= Sample size; M= Mean; SD= Standard deviation.

Results of Three Level Cross-Sectional HLM Analyses

The following section presents the framework that was followed in three-level HLM analyses of Kindergarten, First grade, and Third grade IRT and ARS data sets. Each analysis was comprised of four models starting with a fully unconditional model, followed by a student model with student level predictors in the first level, a classroom model with a classroom level variable (DAP), and finally a school model with a school level predictor (school sector).

I. Fully Unconditional Model

Student level: The model at this level was;

$$\text{Reading (IRT / ARS)}_{ijk} = \pi_{0jk} + e_{ijk}$$

where;

Reading (IRT / ARS) $_{ijk}$: Reading achievement of student *i* in classroom *j* and school *k.*

π_{0jk}: Mean Reading achievement classroom *j* and school *k.*

e_{ijk}: A random student effect. That is, the deviation of student *ijk*'s score from the classroom mean.

Unconditional Level 2 (Classroom level)

$\pi_{0ij} = \beta_{00j} + r_{0jk}$

where;

β_{00k}: Mean reading achievement in school *k.*

r_{0jk}: A random classroom effect. That is, the deviation of classroom *jk*'s score from the school mean.

Unconditional Level 3 (School level)

$\beta_{00j} = \gamma_{000} + u_{00k}$

where,

γ_{000}: Grand mean

u_{00k}: A random school effect. That is, the deviation of school *k*'s from the grand mean.

II. Student Model (Unconditional Level 2 and Level 3)[1]: The next model was the student model in which several student level predictors were added to the first level in order to examine the effects of these variables on students' reading performance and also to explain the variation of reading scores.

Student level

$Reading_{ijk} = \pi_{0jk} + \pi_{1jk}(MALE) + \pi_{2jk}(AGE) + \pi_{3jk}(SES) + \pi_{4jk}(BLACK) + \pi_{5jk}(HISPANIC) + \pi_{6jk}(ASIAN) + \pi_{7jk}(OTHER) + e_{ijk}$

where;

Reading $_{ijk}$: Reading achievement of student *i* in classroom *j* and school *k.*

π_{0jk}: Mean Reading achievement classroom *j* and school *k.*

π_{1jk}: Effect of MALE on reading achievement

π_{2jk}: Effect of AGE on reading achievement

π_{3jk}: Effect of SES on reading achievement

π_{4jk}: Effect of BLACK on reading achievement

π_{5jk}: Effect of HISPANIC on reading achievement

π_{6jk}: Effect of ASIAN on reading achievement

π_{7jk}: Effect of OTHER on reading achievement

e_{ijk}: A random student effect. That is, the deviation of student *ijk*'s score from the classroom mean.

In the following levels, residual variances associated with each student level predictors were fixed due to the fact that effects of student level predictor on reading performance did not varied between classroom within schools and also between schools.

[1] In the Third Grade Date, the predictor AGE was provided as a categorical variable from the ECLS-K. Hence, two dummy coded AGE variables (YOUNGER and OLDER) were included in Third grade models.

Unconditional Level 2 (Classroom level)

$\pi_{0ij} = \beta_{00j} + r_{0jk}$
$\pi_{1jk} = \beta_{10j}$
$\pi_{2jk} = \beta_{20j}$
$\pi_{3jk} = \beta_{30j}$
$\pi_{4jk} = \beta_{40j}$
$\pi_{5jk} = \beta_{50j}$
$\pi_{6jk} = \beta_{60j}$
$\pi_{7jk} = \beta_{70j}$

Unconditional Level 3 (School level)

$\beta_{00j} = \gamma_{000} + u_{00j}$
$\beta_{10j} = \gamma_{100}$
$\beta_{20j} = \gamma_{200}$
$\beta_{30j} = \gamma_{300}$
$\beta_{40j} = \gamma_{400}$
$\beta_{50j} = \gamma_{500}$
$\beta_{60j} = \gamma_{600}$
$\beta_{70j} = \gamma_{700}$

III. Classroom Model (Conditional Level 2 and Unconditional Level 3)

In this model, DAP as a classroom level indicator was added to the model to examine the effect of DAP on reading performance and also the effect of the relationship between DAP and the selected student level predictors on reading achievement, which were the prime interest of this study. Each level model related to the classroom model was illustrated as following;

Student level

$Reading_{ijk} = \pi_{0jk} + \pi_{1jk}(MALE) + \pi_{2jk}(AGE) + \pi_{3jk}(SES) + \pi_{4jk}(BLACK) + \pi_{5jk}(HISPANIC) + \pi_{6jk}(ASIAN) + \pi_{7jk}(OTHER) + e_{ijk}$

where;

Reading $_{ijk}$: Reading achievement of student *i* in classroom *j* and school *k*.

π_{0jk}: Mean Reading achievement classroom *j* and school *k*.

π_{1jk}: Effect of MALE on reading achievement

π_{2jk}: Effect of AGE on reading achievement

π_{3jk}: Effect of SES on reading achievement

π_{4jk}: Effect of BLACK on reading achievement

π_{5jk}: Effect of HISPANIC on reading achievement

π_{6jk}: Effect of ASIAN on reading achievement

π_{7jk}: Effect of OTHER on reading achievement

e_{ijk}: A random student effect. That is, the deviation of student *ijk*'s score from the classroom mean.

In the following levels, residual variances associated with each student level predictors were fixed due to the fact that effects of student level predictor on reading performance did not varied between classroom within schools and also between schools.

Conditional Level 2 (Classroom level)

$$\pi_{0ij} = \beta_{00j} + \beta_{01j}\,(DAP) + r_{0jk}$$

$\pi_{1jk} = \beta_{10j} + \beta_{11j}$ (DAP)
$\pi_{2jk} = \beta_{20j} + \beta_{21j}$ (DAP)
$\pi_{3jk} = \beta_{30j} + \beta_{31j}$ (DAP)
$\pi_{4jk} = \beta_{40j} + \beta_{41j}$ (DAP)
$\pi_{5jk} = \beta_{50j} + \beta_{51j}$ (DAP)
$\pi_{6jk} = \beta_{60j} + \beta_{61j}$ (DAP)
$\pi_{7jk} = \beta_{70j} + \beta_{71j}$ (DAP)

Unconditional Level 3 (School level)

$\beta_{00j} = \gamma_{000} + u_{00j}$
$\beta_{01j} = \gamma_{010}$
$\beta_{10j} = \gamma_{100}$
$\beta_{11j} = \gamma_{110}$
$\beta_{20j} = \gamma_{200}$
$\beta_{21j} = \gamma_{210}$
$\beta_{30j} = \gamma_{300}$
$\beta_{31j} = \gamma_{310}$
$\beta_{40j} = \gamma_{400}$
$\beta_{41j} = \gamma_{410}$
$\beta_{50j} = \gamma_{500}$
$\beta_{51j} = \gamma_{510}$
$\beta_{60j} = \gamma_{600}$
$\beta_{61j} = \gamma_{610}$

$\beta_{70j} = \gamma_{700}$
$\beta_{71j} = \gamma_{710}$

IV. School Model (Conditional Level 2 and Level 3)

In this model, PUBLIC, indicating school type (sector), was added to the model to examine the effect of school sector on reading performance. Each level model related to the classroom model was illustrated as following;

Student level

$\text{Reading}_{ijk}=\pi_{0jk}+ \pi_{1jk}(\text{MALE})+ \pi_{2jk}(\text{AGE})+ \pi_{3jk}(\text{SES})+ \pi_{4jk}(\text{BLACK})+ \pi_{5jk}(\text{HISPANIC})+ \pi_{6jk}(\text{ASIAN})+ \pi_{7jk}(\text{OTHER})+e_{ijk}$

where;

Reading_{ijk}: Reading achievement of student *i* in classroom *j* and school *k.*

π_{0jk}: Mean Reading achievement classroom *j* and school *k.*

π_{1jk}: Effect of MALE on reading achievement

π_{2jk}: Effect of AGE on reading achievement

π_{3jk}: Effect of SES on reading achievement

π_{4jk}: Effect of BLACK on reading achievement

π_{5jk}: Effect of HISPANIC on reading achievement

π_{6jk}: Effect of ASIAN on reading achievement

π_{7jk}: Effect of OTHER on reading achievement

e_{ijk}: A random student effect. That is, the deviation of student *ijk*'s score from the classroom mean.

In the following levels, residual variances associated with each student level predictors were fixed due to the fact that effects of student level predictor on reading performance did not varied between classroom within schools and also between schools.

Conditional Level 2 (Classroom level)

$\pi_{0ij} = \beta_{00j} + \beta_{01j}$ (DAP) $+ r_{0jk}$

$\pi_{1jk} = \beta_{10j}$
$\pi_{2jk} = \beta_{20j}$
$\pi_{3jk} = \beta_{30j}$
$\pi_{4jk} = \beta_{40j}$
$\pi_{5jk} = \beta_{50j}$
$\pi_{6jk} = \beta_{60j}$
$\pi_{7jk} = \beta_{70j}$

Conditional Level 3 (School level)

$\beta_{00j} = \gamma_{000} + \gamma_{001}$(PUBLIC) $+ u_{00j}$
$\beta_{01j} = \gamma_{010} + \gamma_{011}$(PUBLIC)
$\beta_{10j} = \gamma_{100}$
$\beta_{20j} = \gamma_{200}$
$\beta_{30j} = \gamma_{300}$
$\beta_{40j} = \gamma_{400}$
$\beta_{50j} = \gamma_{500}$
$\beta_{60j} = \gamma_{600}$
$\beta_{70j} = \gamma_{700}$

Kindergarten IRT Data

The results of the fully unconditional model, presented in Table 13, showed that grand mean of student reading scores was 32.19 (γ_{000}). In other words, the weighted least squares estimate for the grand mean reading achievement at the Kindergarten level was estimated as 32.19 (t=180.51, p<.001). It had a standard error of .18 and a 95% confidence interval was computed by;

$$95\%CI(\gamma_{000}) = \gamma \pm 1.96\ (V\gamma_{00})^{1/2}$$

where $V\gamma_{00}$ is the estimated sampling variance of γ_{000}.
The calculation yielded a range of

$$32.19 \pm 1.96\ (.18) = (31.84, 32.54).$$

Table 13
Results of the Three-level-Analysis of the Kindergarten IRT Data (Fully Unconditional Model)

Fixed Effects	*Coefficient*	*se*	*t- ratio*
Classroom model for student level intercept, π_{0jk}			
School model for classroom level intercept, β_{00k}			
Intercept, γ_{000}	32.19*	.18	180.50

Random Effects	Variance	*df*	χ^2	*p-value*
School mean, u_{00k}	19.14	940	2957.65	.001
Classroom mean, r_{0jk}	9.30	1952	3632.73	.001
Students, e_{ijk}	9.01			

Note: se= standard error of estimate; df=degrees of freedom; χ^2*=chi-square statistic;* * $p<.05$.

The estimates of the variance components were given in the second part of Table 13. The chi-square statistics associated with these variance components indicated significant variations among students within classrooms, among classrooms with schools and also between schools. Specifically, the variance components were estimated as 9.01 (σ^2) among students over classrooms, 9.30 (τ_π) between classrooms within schools, and 19.14 (τ_β) between schools. To estimate of the percentage of variance in the reading achievement at each level, the following equations were used;

% variance among students within classrooms: $\sigma^2 / (\sigma^2 + \tau_\pi + \tau_\beta)$,

% variance among classrooms within schools: $\tau_\pi / (\sigma^2 + \tau_\pi + \tau_\beta)$,

% variance between schools: $\tau_\beta / (\sigma^2 + \tau_\pi + \tau_\beta)$,

Substituting the estimates for each of these variance components into equation, revealed that 24% of the variance laid between students, 25% of the variance laid between classrooms within schools, and 51% of the total variance laid between schools. In an effort to estimate the magnitude of the variation between schools in their mean reading achievement levels, it would be useful to calculate the plausible values range (Raudenbush & Bryk, 2002) for these means. It was expected that 95% of the school means to fall within the range;

$$\hat{\gamma} \pm 1.96\,(\hat{\tau}_{00})^{1/2},$$

which yields

$$32.19 \pm 1.96\ (19.14)^{1/2} = (23.62, 40.76)$$

This range is a substantial range in average reading achievement levels among schools in the Kindergarten IRT data. The result of the test for whether all schools have the same mean was found to be significant at .05 level (χ^2=2957.65, p<.001). This means that there is a significant variation between schools in terms of reading achievement.

The student level model presented in Table 14 displayed that the average school reading IRT achievement mean (grand mean) was 32.18 (γ_{000}) (p<.001) for the typical/reference student (Female=0, White=0, average respect to the AGE and SES). Results from the student model's estimated fixed effects indicated that all effects of selected student level predictors were statistically significant on the IRT performance (p<.001). Specifically, male students' reading achievement on the IRT scale was 1.90 (γ_{100}) points lower than female students. The relationship between AGE and the reading performance was estimated as .35 (γ_{200}), indicating that a student, who was older one standard deviation, scored .35 points greater than younger student. The effect of SES on the estimated reading IRT achievement was found to be statistically significant.

The estimated effect of SES, (γ_{300}=3.31), indicated that a student with one standard deviation higher SES level scored 3.31 points higher than others. The effect of BLACK on the reading performance was –2.13 (γ_{400}), indicating that an African-American student scored 2.13 points lower than his or her Caucasian counterparts within school *k*. The effect of HISPANIC on the reading performance was – 1.59 (γ_{500}), indicating that a Hispanic student scored 1.59 points lower than Caucasians. However, the effect of ASIAN on the reading performance was found to be 2.20 (γ_{600}), indicating that an Asian student scored 2.20 points greater than his or her Caucasian peer within school *k*. Finally, the effect of OTHER on the reading performance was found to be –1.22 (γ_{700}), indicating that a student who was in the other race group rather than white, black, Hispanic, or Asian, scored 1.22 points lower than Caucasians within school *k*.

Table 14
Results of the Three-level-Analysis of the Kindergarten IRT Data (Student Model)

Fixed Effects	*Coefficient*	*se*	*t- ratio*
Classroom model for student level intercept, π_{0jk}			
School model for classroom level intercept, β_{00k}			
Intercept, γ_{000}	32.18*	.17	180.15
Classroom model for student level effect of MALE, π_{1jk}			
School model for classroom level intercept, β_{10k}			
Intercept, γ_{100}	-1.90*	.16	-11.78
Classroom model for student level effect of AGE, π_{2jk}			
School model for classroom level intercept, β_{20k}			
Intercept, γ_{200}	.35*	.02	15.37
Classroom model for student level effect of SES, π_{3jk}			
School model for classroom level intercept, β_{30k}			
Intercept, γ_{300}	3.31*	.14	22.46
Classroom model for student level effect of BLACK, π_{4jk}			
School model for classroom level intercept, β_{40k}			
Intercept, γ_{400}	-2.13*	.35	-5.99
Classroom model for student level effect of HISPANIC, π_{5jk}			
School model for classroom level intercept, β_{50k}			
Intercept, γ_{500}	-1.59*	.32	-4.89
Classroom model for student level effect of ASIAN, π_{6jk}			
School model for classroom level intercept, β_{60k}			
Intercept, γ_{600}	2.20*	.55	3.97
Classroom model for student level effect of OTHER, π_{7jk}			
School model for classroom level intercept, β_{70k}			
Intercept, γ_{700}	-1.22*	.47	-2.58

Random Effects	Variance	*df*	χ^2	*p-value*
Schools, u_{00k}	19.08	940	2924.89	.001
Classrooms, r_{0jk}	9.22	1952	3522.31	.001
Students, e_{ijk}	8.15			

*Note: se= standard error of estimate; df=degrees of freedom; χ^2=chi-square statistic; * p<.05.*

Estimated variances and related chi square statistics from three level of the student model were presented at the end of the Table 14. These results suggested that even though residual parameter variances in all levels decreased slightly, they still remained unexplained in π_{0jk} (8.15), β_{00j} (9.22), and γ_{00k} (19.08).

Results of the classroom model indicated that there was no evidence of a significant effect of DAP on the reading performance (γ_{010}=.06, p>.500). Additionally, the strengths of associations between reading achievement and student level predictors in the model did not differ when considering the effect of DAP. In other words, the effects of interactions between selected student characteristics and DAP were not found to be significant on the IRT reading achievement scores at .05 level.

The estimated variances of the intercepts were slightly smaller than the previous student model, without controlling for DAP. Therefore, significant variation in the intercepts, including the variance of reading performance between classroom within schools (r_{0jk} =9.15, χ^2=3418.53, df=1951, p<.001) and the variance of reading performance between schools (u_{00k} =19.03, χ^2=2814.92, df=940, p<.001) remained unexplained even after controlling for DAP.

The results of the final (school) model are presented in Table 16. The results of the fixed effects indicated that private schools had significantly higher mean reading achievement than did public schools (γ_{001}= -4.46, p<.001; coded as Public=1, Private=0). In terms of interaction between the effect of DAP and the effect of PUBLIC, it was observed that public schools had significantly larger DAP slopes, on average, than did private schools (γ_{011}=.45, p<.05). It means that public and private schools differed with respect to their average DAP effect on student predictors-reading IRT achievement relationships within schools.

Table 15

Results of the Three-level-Analysis of the Kindergarten IRT Data (Classroom Model)

Fixed Effects	*Coefficient*	*se*	*t- ratio*
Classroom model for student level intercept, π_{0jk}			
School model for classroom level intercept, β_{00k}			
Intercept, γ_{000}	32.18*	.17	180.29
School model for classroom level effect of DAP, β_{01k}			
Intercept, γ_{010}	.06	.05	1.25
Classroom model for student level effect of MALE, π_{1jk}			
School model for classroom level intercept, β_{10k}			
Intercept, γ_{100}	-1.91*	.15	-12.55
School model for classroom level effect of DAP, β_{11k}			
Intercept, γ_{110}	-.00	.06	-.04
Classroom model for student level effect of AGE, π_{2jk}			
School model for classroom level intercept, β_{20k}			
Intercept, γ_{200}	.35*	.02	17.02
School model for classroom level effect of DAP, β_{21k}			
Intercept, γ_{210}	.01	.00	1.45
Classroom model for student level effect of SES, π_{3jk}			
School model for classroom level intercept, β_{30k}			
Intercept, γ_{300}	3.29*	.12	26.00
School model for classroom level effect of DAP, β_{31k}			
Intercept, γ_{310}	.08	.04	1.71
Classroom model for student level effect of BLACK, π_{4jk}			
School model for classroom level intercept, β_{40k}			
Intercept, γ_{400}	-2.13*	.31	-6.69
School model for classroom level effect of DAP, β_{41k}			
Intercept, γ_{410}	.06	.12	.47
Classroom model for student level effect of HISPANIC, π_{5jk}			
School model for classroom level intercept, β_{50k}			
Intercept, γ_{500}	-1.58*	.29	-5.47
School model for classroom level effect of DAP, β_{51k}			
Intercept, γ_{510}	-.08	.11	-.75

Table 15 (continued).

Fixed Effects	*Coefficient*	*se*	*t- ratio*
Classroom model for student level effect of ASIAN, π_{6jk}			
School model for classroom level intercept, β_{60k}			
Intercept, γ_{600}	2.31*	.53	4.31
School model for classroom level effect of DAP, β_{61k}			
Intercept, γ_{610}	-.22	.20	-1.07
Classroom model for student level effect of OTHER, π_{7jk}			
School model for classroom level intercept, β_{70k}			
Intercept, γ_{700}	-1.25*	.43	-2.89
School model for classroom level effect of DAP, β_{71k}			
Intercept, γ_{710}	.27	.17	1.59

Random Effects	Variance	*df*	χ^2	*p-value*
Schools, u_{00k}	19.03	940	2814.92	.001
Classrooms, r_{0jk}	9.15	1951	3418.53	.001
Students, e_{ijk}	8.11			

Note: se= standard error of estimate; df=degrees of freedom; χ^2*=chi-square statistic;* * $p<.05$.

Examination of the variance components placed at the end of the Table 16 indicated that significant variation in the intercepts, including the variance of reading performance between classroom within schools (r_{0jk} =9.12, χ^2=3398.67, *df*=1951, p<.001) and the variance of reading performance between schools (u_{00k} =15.81, χ^2=2653.31, *df*=939, p<.001) remained unexplained even after controlling for PUBLIC.

Finally, a proportion of reduction in variance for each of the random coefficients, herein intercepts, was computed using;

$$\text{Proportion of variance explained in } y_{000} =$$

$$[\hat{\tau}_{000}\text{ (null model)} - \hat{\tau}_{000}\text{ (fitted model)}] / \hat{\tau}_{000}\text{ (random regression)}$$

where, $\hat{\tau}_{000}$ (null model) refers to the fully unconditional (null) variance of the intercept, which was 19.14 and the $\hat{\tau}_{000}$ (fitted model) refers to the residual variance of the final (school model), which was estimated as 15.81. Therefore, the proportion variance explained in β_{000} was computed as = [19.14-15.81] / 19.14 =.17, indicating that 17% of the parameter variation in mean reading achievement between schools was explained by adding all variables.

Table 16

Results of the Three-level-Analysis of the Kindergarten IRT Data (School Model)

Fixed Effects	*Coefficient*	*se*	*t- ratio*
Classroom model for student level intercept, π_{0jk}			
School model for classroom level intercept, β_{00k}			
Intercept, γ_{000}	32.35*	.17	187.17
PUBLIC, γ_{001}	-4.46*	.46	-9.63
School model for classroom level effect of DAP, β_{01k}			
Intercept, γ_{010}	.09	.05	1.75
PUBLIC, γ_{011}	.45*	.13	3.31
Classroom model for student level effect of MALE, π_{1jk}			
School model for classroom level intercept, β_{10k}			
Intercept, γ_{100}	-1.90*	.16	-11.78
Classroom model for student level effect of AGE, π_{2jk}			
School model for classroom level intercept, β_{20k}			
Intercept, γ_{200}	.35*	.02	15.37
Classroom model for student level effect of SES, π_{3jk}			
School model for classroom level intercept, β_{30k}			
Intercept, γ_{300}	3.31*	.14	22.46
Classroom model for student level effect of BLACK, π_{4jk}			
School model for classroom level intercept, β_{40k}			
Intercept, γ_{400}	-2.13*	.35	-5.99
Classroom model for student level effect of HISPANIC, π_{5jk}			
School model for classroom level intercept, β_{50k}			
Intercept, γ_{500}	-1.59*	.32	-4.89
Classroom model for student level effect of ASIAN, π_{6jk}			
School model for classroom level intercept, β_{60k}			
Intercept, γ_{600}	2.20*	.55	3.97
Classroom model for student level effect of OTHER, π_{7jk}			
School model for classroom level intercept, β_{70k}			
Intercept, γ_{700}	-1.22*	.47	-2.58

Random Effects	Variance	*df*	χ^2	*p-value*
Schools, u_{00k}	15.81	939	2053.31	.001
Classrooms, r_{0jk}	9.12	1951	3398.67	.001
Students, e_{ijk}	8.04			

*Note: se= standard error of estimate; df=degrees of freedom; χ^2=chi-square statistic; * p<.05.*

Kindergarten ARS Data

The results of the fully unconditional model, as shown in Table 17, indicated that grand mean of student reading scores was 3.35 (γ_{000}). That is, the weighted least squares estimate for the grand mean reading achievement at the Kindergarten level was estimated as 3.35 (t=259.98, p<.001). It had a standard error of .01 and a 95% confidence interval was computed by;

$$95\%\mathrm{CI}(\gamma_{000}) = \gamma \pm 1.96\ (V\gamma_{00})^{1/2}$$

where $V\gamma_{00}$ is the estimated sampling variance of γ_{000}.

The calculation yielded a range of

$$3.35 \pm 1.96\ (.01) = (3.33,\ 3.37).$$

Table 17

Results of the Three-level-Analysis of the Kindergarten ARS Data (Fully Unconditional Model)

Fixed Effects	*Coefficient*	*se*	*t- ratio*
Classroom model for student level intercept, π_{0jk}			
School model for classroom level intercept, β_{00k}			
Intercept, γ_{000}	3.35*	.01	259.98

Random Effects	Variance	*df*	χ^2	*p-value*
School mean, u_{00k}	.06	971	1732.53	.001
Classroom mean, r_{0jk}	.15	2070	5735.26	.001
Students, e_{ijk}	.05			

Note: se= standard error of estimate; df=degrees of freedom; χ^2=chi-square statistic; * p<. 05.

The estimates of the variance components were presented in the second part of Table 17. The chi-square statistics associated with these variance components indicated significant variations among students within classrooms, among classrooms with schools and also between schools. Specifically, the variance components were estimated as .05 (σ^2) among students over classrooms, .15 (τ_π) between classrooms within schools, and .06 (τ_β) between schools. To estimate of the percentage of variance in the reading achievement at each level, the following equations were used;

% variance among students within classrooms: $\sigma^2 / (\sigma^2 + \tau_\pi + \tau_\beta)$,

% variance among classrooms within schools: $\tau_\pi / (\sigma^2 + \tau_\pi + \tau_\beta)$,

% variance between schools: $\tau_\beta / (\sigma^2 + \tau_\pi + \tau_\beta)$,

Substituting the estimates for each of these variance components into equations, it was estimated that 19% of the variance laid between students, 58% of the variance laid between classrooms within schools, and 23% of the total variance laid between schools. Unlike the results of variance components in the Kindergarten IRT Data, most of the variance was at the classroom level.

The estimate of the between school variability (the variance of school means around the grand-mean) is .06. Therefore, 95% of the school means fall within the following range:

$$3.35 \pm 1.96\,(.06)^{1/2} = (2.87, 3.83)$$

The result of the test for whether all schools have the same mean was found to be significant at .05 level (χ^2=1732.53, p<.001). That is to say, there is a significant variation between schools in terms of reading achievement in ARS scale.

The student level model is illustrated in Table 18 and the average school reading ARS achievement mean (grand mean) was 3.34 (γ_{000}) (p<.001) for the typical student (Female=0, White=0, average respect to the AGE and SES). Results from the student model's estimated fixed effects indicated that all effects of selected student level predictors were statistically significant on the reading ARS performance (p<.001). Specifically, male students' reading achievement on the ARS scale was .20 (γ_{100}) points lower than female students. The relationship between AGE and the reading performance was estimated as .03 (γ_{200}), indicating that a student, who was older one standard deviation, scored .03 points greater than younger student. The effect of SES on the reading ARS achievement was found to be statistically significant. The estimated effect of SES, (γ_{300} =.27), indicated that a student with one standard deviation higher SES level scored .27 points higher than others. The effect of BLACK on the reading performance was –.14 (γ_{400}), referring that an African-American student scored .14 points lower than his or her counterparts who were Caucasian within school k. The effect of HISPANIC on the reading achievement score was – .19 (γ_{500}), indicating that a Hispanic student scored .19 points lower than Caucasians. The effect of ASIAN on the reading performance

Table 18

Results of the Three-level-Analysis of the Kindergarten ARS Data (Student Model)

Fixed Effects	*Coefficient*	*se*	*t- ratio*
Classroom model for student level intercept, π_{0jk}			
School model for classroom level intercept, β_{00k}			
Intercept, γ_{000}	3.34*	.01	259.53
Classroom model for student level effect of MALE, π_{1jk}			
School model for classroom level intercept, β_{10k}			
Intercept, γ_{100}	-.20*	.01	-16.14
Classroom model for student level effect of AGE, π_{2jk}			
School model for classroom level intercept, β_{20k}			
Intercept, γ_{200}	.03*	.00	18.99
Classroom model for student level effect of SES, π_{3jk}			
School model for classroom level intercept, β_{30k}			
Intercept, γ_{300}	.27*	.01	23.64
Classroom model for student level effect of BLACK, π_{4jk}			
School model for classroom level intercept, β_{40k}			
Intercept, γ_{400}	-.14*	.03	-4.71
Classroom model for student level effect of HISPANIC, π_{5jk}			
School model for classroom level intercept, β_{50k}			
Intercept, γ_{500}	-.19*	.02	-7.26
Classroom model for student level effect of ASIAN, π_{6jk}			
School model for classroom level intercept, β_{60k}			
Intercept, γ_{600}	-.03*	.03	-1.11
Classroom model for student level effect of OTHER, π_{7jk}			
School model for classroom level intercept, β_{70k}			
Intercept, γ_{700}	-.11*	.03	-3.34

Random Effects	Variance	*df*	χ^2	*p-value*
School mean, u_{00k}	.06	971	1714.69	.001
Classroom mean, r_{0jk}	.16	2070	6243.38	.001
Students, e_{ijk}	.04			

*Note: se= standard error of estimate; df=degrees of freedom; χ^2=chi-square statistic; * p<.05.*

was found to be -.03 (γ_{600}), referring that an Asian student scored .03 points lower than his or her Caucasian counterparts within school *k*. Finally, the effect of OTHER on the reading achievement score was found to be –.11 (γ_{700}), indicating that a student who was in the other race group rather than white, black, Hispanic, or Asian, scored .11 points lower than Caucasian students within school *k*.

Estimated variances and related chi square statistics from three levels of the student model were presented at the end of the Table 18. These results suggested that even though residual parameter variances in all levels decreased slightly, they still remained unexplained in π_{0jk} (.04), β_{00j} (.16), and γ_{00k} (.06).

The classroom level variable of DAP was added to the model in the next step to examine the effect of DAP on reading performance and also the effect of the relationship between DAP and the selected student level predictors on reading achievement. The findings at this model revealed that DAP had a significant effect on the reading score (γ_{010}= .03, t= 6.89, p<.05) (See Table 19). Additionally, the strengths of associations between reading achievement and most of the student level predictors in the model did not differ when considering the effect of DAP, except for the interaction effect between SES and DAP on performance. The effect of interaction between selected student characteristics and DAP were not found to be significant on the ARS reading achievement at .05 level. The effect of DAP on the effect of SES was statistically significant and positive in direction (γ_{310}= .01, t= 3.04, p< .01). The reading performance for students with higher SES levels increased significantly as a function of DAP.

Table 19

Results of the Three-level-Analysis of the Kindergarten ARS Data (Classroom Model)

Fixed Effects	*Coefficient*	*se*	*t- ratio*
Classroom model for student level intercept, π_{0jk}			
School model for classroom level intercept, β_{00k}			
Intercept, γ_{000}	3.34*	.01	263.18
School model for classroom level effect of DAP, β_{01k}			
Intercept, γ_{010}	.03*	.00	6.89
Classroom model for student level effect of MALE, π_{1jk}			
School model for classroom level intercept, β_{10k}			
Intercept, γ_{100}	-.20*	.01	-18.13
School model for classroom level effect of DAP, β_{11k}			
Intercept, γ_{110}	-.00	.00	-1.73
Classroom model for student level effect of AGE, π_{2jk}			
School model for classroom level intercept, β_{20k}			
Intercept, γ_{200}	.03*	.00	21.43
School model for classroom level effect of DAP, β_{21k}			
Intercept, γ_{210}	-.00	.00	-.00
Classroom model for student level effect of SES, π_{3jk}			
School model for classroom level intercept, β_{30k}			
Intercept, γ_{300}	.27*	.00	29.18
School model for classroom level effect of DAP, β_{31k}			
Intercept, γ_{310}	.01*	.00	3.04
Classroom model for student level effect of BLACK, π_{4jk}			
School model for classroom level intercept, β_{40k}			
Intercept, γ_{400}	-.14*	.02	-5.96
School model for classroom level effect of DAP, β_{41k}			
Intercept, γ_{410}	.00	.00	.21
Classroom model for student level effect of HISPANIC, π_{5jk}			
School model for classroom level intercept, β_{50k}			
Intercept, γ_{500}	-.18*	.02	-8.84
School model for classroom level effect of DAP, β_{51k}			
Intercept, γ_{510}	-.01	.00	-1.31

Table 19 (continued).

Fixed Effects	*Coefficient*	*se*	*t- ratio*
Classroom model for student level effect of ASIAN, π_{6jk}			
School model for classroom level intercept, β_{60k}			
Intercept, γ_{600}	-.03*	.03	-.94
School model for classroom level effect of DAP, β_{61k}			
Intercept, γ_{610}	-.00	.01	-.22
Classroom model for student level effect of OTHER, π_{7jk}			
School model for classroom level intercept, β_{70k}			
Intercept, γ_{700}	-.11*	.03	-3.29
School model for classroom level effect of DAP, β_{71k}			
Intercept, γ_{710}	.00	.01	.01

Random Effects	Variance	*df*	χ^2	*p-value*
School mean, u_{00k}	.05	971	934.30	.001
Classroom mean, r_{0jk}	.15	2069	6204.20	.001
Students, e_{ijk}	.04			

*Note: se= standard error of estimate; df=degrees of freedom; χ^2=chi-square statistic; * p<.05.*

The estimated variances of the intercepts were slightly smaller than the previous student model, without controlling for DAP. Therefore, significant variation in the intercepts, including the variance of reading performance between classroom within schools (r_{0jk} =.15, χ^2=6204.20, *df*=1098, p<.001) and the variance of reading performance between schools (u_{00k} =.05, χ^2=934.30, *df*=820, p<.001) remained unexplained even after controlling for DAP.

School sector variable of PUBLIC was added to the school model, to examine the effect of school sector on reading performance. Results of the final (school) model are presented in Table 20. The findings of the fixed effects indicated that private schools had higher mean reading achievement than did public schools (γ_{001}= –.24, p>.05; coded as Public=1, Private=0), but this effect was not statistically significant as indicated.

Table 20

Results of the Three-level-Analysis of the Kindergarten ARS Data (School Model)

Fixed Effects	*Coefficient*	*se*	*t- ratio*
Classroom model for student level intercept, π_{0jk}			
School model for classroom level intercept, β_{00k}			
Intercept, γ_{000}	3.36*	.01	245.38
PUBLIC, γ_{001}	-.24	.03	-1.69
School model for classroom level effect of DAP, β_{01k}			
Intercept, γ_{010}	.03*	.00	4.93
PUBLIC, γ_{011}	.01	.01	-1.03
Classroom model for student level effect of MALE, π_{1jk}			
School model for classroom level intercept, β_{10k}			
Intercept, γ_{100}	-.20*	.01	-1.54
Classroom model for student level effect of AGE, π_{2jk}			
School model for classroom level intercept, β_{20k}			
Intercept, γ_{200}	.03*	.00	18.01
Classroom model for student level effect of SES, π_{3jk}			
School model for classroom level intercept, β_{30k}			
Intercept, γ_{300}	.27*	.00	29.18
School model for classroom level effect of DAP, β_{31k}			
Intercept, γ_{310}	.01*	.00	3.21
PUBLIC, γ_{311}	.00	.01	0.24
Classroom model for student level effect of BLACK, π_{4jk}			
School model for classroom level intercept, β_{40k}			
Intercept, γ_{400}	-.14*	.02	4.52
Classroom model for student level effect of HISPANIC, π_{5jk}			
School model for classroom level intercept, β_{50k}			
Intercept, γ_{500}	-.19*	.02	-7.13
Classroom model for student level effect of ASIAN, π_{6jk}			
School model for classroom level intercept, β_{60k}			
Intercept, γ_{600}	-.03*	.03	-2.81
Classroom model for student level effect of OTHER, π_{7jk}			
School model for classroom level intercept, β_{70k}			
Intercept, γ_{700}	-.11	.03	1.72

Random Effects	Variance	*df*	χ^2	*p-value*
School mean, u_{00k}	.05	970	1639.34	.001
Classroom mean, r_{0jk}	.15	2069	6333.82	.001
Students, e_{ijk}	.04			

*Note: se= standard error of estimate; df=degrees of freedom; χ^2=chi-square statistic; * p<.05.*

There was no evidence of a significant effect of interaction between PUBLIC and DAP on the reading achievement scores (γ_{011}=.01, p>.05). It means that public and private schools did not differ with respect to their average DAP-reading ARS achievement relationship with schools. Similarly, the effect of PUBLIC on the effect of SES-reading achievement slope (γ_{311}=.00, p>.05) was not statistically significant at .05 level, indicating that public and private schools did not vary with respect to their average DAP effect on SES-reading ARS achievement relationship within schools.

Examination of the variance components (see Table 20) indicated that significant variation in the intercepts, including the variance of reading performance between classrooms within schools (r_{0jk} =.15, χ^2=6333.82, *df*=2069, p<.001) and the variance of reading performance between schools (u_{00k} =.05, χ^2=1639.34, *df*=970, p<.001) remained unexplained even after controlling for PUBLIC. The proportion of variance explained in y_{000} was estimated as = [.06-.05] / .06 =.17, indicating that 17% of the parameter variation in mean reading achievement between schools was explained by adding all variables.

First Grade IRT Data

The findings regarding the fully unconditional model for the first grade IRT data, presented in Table 21, indicated that grand mean of student reading scores was 56.86 (γ_{000}) in first grade. In other words, the weighted least squares estimate for the grand mean reading achievement at the first grade level was estimated as 56.86 (t=236.11, p<.001). It had a standard error of .24 and a 95% confidence interval was;

$$56.86 \pm 1.96\ (.24) = (56.39,\ 57.33).$$

The chi-square statistics associated with these variance components indicated significant variations among students within classrooms, among classrooms with schools and also between schools. Specifically, the variance components were estimated as 15.51 (σ^2) among students over classrooms, 27.26 (τ_π) between classrooms within schools, and 38.68 (τ_β) between schools.

Table 24
Results of the Three-level-Analysis of the First Grade IRT Data (Fully Unconditional Model)

Fixed Effects	*Coefficient*	*se*	*t- ratio*
Classroom model for student level intercept, π_{0jk}			
School model for classroom level intercept, β_{00k}			
Intercept, γ_{000}	56.86*	.24	236.11

Random Effects	Variance	*df*	χ^2	*p-value*
School mean, u_{00k}	38.68	1214	3056.56	.001
Classroom mean, r_{0jk}	27.26	2041	4285.40	.001
Students, e_{ijk}	15.51			

*Note: se= standard error of estimate; df=degrees of freedom; χ^2=chi-square statistic; * p<. 05.*

To estimate of the percentage of variance in the reading achievement at each level, the following equations were used;

% variance among students within classrooms: $\sigma^2 / (\sigma^2 + \tau_\pi + \tau_\beta)$,

% variance among classrooms within schools: $\tau_\pi / (\sigma^2 + \tau_\pi + \tau_\beta)$,

% variance between schools: $\tau_\beta / (\sigma^2 + \tau_\pi + \tau_\beta)$,

Substituting the estimates for each of these variance components into equations, revealed that 19% of the variance laid between students, 33% of the variance laid between classrooms within schools, and 48% of the total variance laid between schools, respectively. As can be seen, most of the variance was at the school level.

The estimate of the between school variability (the variance of school means around the grand-mean) was 38.68. Therefore, 95% of the school means fall within the following range:

$$56.86 \pm 1.96\,(38.68)^{1/2} = (44.67, 69.05)$$

The result of the test for whether all schools have the same mean was found to be significant at .05 level (χ^2=3056.56, p<.001), indicating that there is a significant variation between schools in terms of first grade reading achievement in IRT scale.

In the student model, a total of seven student level predictors were added to the student level in order to examine the effects of these variables on students' reading performance and also to explain the variation of reading scores.

The student level model presented in Table 22 displayed that the average school reading IRT achievement mean (grand mean) was 56.85 (γ_{000}) (p<.001) for the typical student (Female=0, White=0, average respect to the AGE and SES). Results from the student model's estimated fixed effects indicated that all effects of selected student level predictors, except OTHER, were statistically significant on the reading IRT scores (p<.001). Specifically, male students' reading achievement on the IRT scale was 2.09 (γ_{100}) points lower than female students. The relationship between AGE and the reading performance was estimated as .16 (γ_{200}), indicating that a student, who was older one standard deviation, scored .16 points greater than younger student. The effect of SES on reading IRT achievement scores was found to be statistically significant. The estimated effect of SES, (γ_{300} =3.59), indicated that a student with one standard deviation higher SES level scored 3.59 points higher than others. The effect of BLACK on the reading performance was –2.37 (γ_{400}), indicating that an African-American student scored 2.37 points lower than his or her Caucasian counterparts within school *k*.

The effect of HISPANIC on the reading performance was found to be – 1.47 (γ_{500}), indicating that a Hispanic student were outscored 1.47 points by Caucasians. The effect of ASIAN on reading achievement scores was 2.06 (γ_{600}), referring to a 2.06 points advantage for Asian students over their Caucasian peers within school *k*. Finally, the effect of OTHER on the reading performance was –.87 (γ_{700}), indicating that a student who was in the other race group rather than white, black, Hispanic, or Asian, scored .87 points lower than his or her Caucasian counterparts within school *k*, but this effect was not significant.

Table 22

Results of the Three-level-Analysis of the First Grade IRT Data (Student Model)

Fixed Effects	*Coefficient*	*se*	*t- ratio*
Classroom model for student level intercept, π_{0jk}			
School model for classroom level intercept, β_{00k}			
Intercept, γ_{000}	56.85*	.24	235.69
Classroom model for student level effect of MALE, π_{1jk}			
School model for classroom level intercept, β_{10k}			
Intercept, γ_{100}	-2.09*	.24	-8.65
Classroom model for student level effect of AGE, π_{2jk}			
School model for classroom level intercept, β_{20k}			
Intercept, γ_{200}	.16*	.03	5.05
Classroom model for student level effect of SES, π_{3jk}			
School model for classroom level intercept, β_{30k}			
Intercept, γ_{300}	3.59*	.20	17.21
Classroom model for student level effect of BLACK, π_{4jk}			
School model for classroom level intercept, β_{40k}			
Intercept, γ_{400}	-2.37*	.55	-4.46
Classroom model for student level effect of HISPANIC, π_{5jk}			
School model for classroom level intercept, β_{50k}			
Intercept, γ_{500}	-1.47*	.47	-2.54
Classroom model for student level effect of ASIAN, π_{6jk}			
School model for classroom level intercept, β_{60k}			
Intercept, γ_{600}	2.06*	.80	2.36
Classroom model for student level effect of OTHER, π_{7jk}			
School model for classroom level intercept, β_{70k}			
Intercept, γ_{700}	-.87	.74	-1.23

Random Effects	Variance	*df*	χ^2	*p-value*
Schools, u_{00k}	38.81	1214	3057.98	.001
Classrooms, r_{0jk}	29.17	2041	4399.59	.001
Students, e_{ijk}	14.65			

*Note: se= standard error of estimate; df=degrees of freedom; χ^2=chi-square statistic; * p<.05.*

Estimated variances and related chi square statistics from three level of the student model were presented at the end of the Table 22. These results suggested that even though residual parameter variances in all levels decreased slightly, they still remained unexplained in π_{0jk} (14.65), β_{00j} (29.17), and γ_{00k} (38.31).

In the classroom model, DAP as a classroom level indicator was added to the model to examine the effect of DAP on reading performance and also the effect of the relationship between DAP and the selected student level predictors on reading achievement. Results of the classroom model indicated that DAP did not have a significant effect on the reading performance (γ_{010}= –.08, t= –.96, p>.05) (See Table 23). Similarly, the strengths of relations between reading achievement and all of the student level predictors in the model did not differ when considering the effect of DAP.

The estimated variances of the intercepts were slightly smaller than the previous student model, without controlling for DAP. Therefore, significant variation in the intercepts, including the variance of reading performance between classroom within schools (r_{0jk} =29.12, χ^2=4384.18, df=2040, p<.001) and the variance of reading performance between schools (u_{00k} =38.61, χ^2=2916.74, df=1214, p<.001) remained unexplained even after controlling for DAP.

Table 23

Results of the Three-level-Analysis of the First Grade IRT Data (Classroom Model)

Fixed Effects	*Coefficient*	*se*	*t- ratio*
Classroom model for student level intercept, π_{0jk}			
School model for classroom level intercept, β_{00k}			
Intercept, γ_{000}	56.84*	.24	235.78
School model for classroom level effect of DAP, β_{01k}			
Intercept, γ_{010}	-.08	.09	-.96
Classroom model for student level effect of MALE, π_{1jk}			
School model for classroom level intercept, β_{10k}			
Intercept, γ_{100}	-2.13*	.24	-8.74
School model for classroom level effect of DAP, β_{11k}			
Intercept, γ_{110}	-.14	.12	-1.18
Classroom model for student level effect of AGE, π_{2jk}			
School model for classroom level intercept, β_{20k}			
Intercept, γ_{200}	.16*	.03	5.01
School model for classroom level effect of DAP, β_{21k}			
Intercept, γ_{210}	.00	.01	.34
Classroom model for student level effect of SES, π_{3jk}			
School model for classroom level intercept, β_{30k}			
Intercept, γ_{300}	3.62*	.20	17.49
School model for classroom level effect of DAP, β_{31k}			
Intercept, γ_{310}	.12	.10	1.23
Classroom model for student level effect of BLACK, π_{4jk}			
School model for classroom level intercept, β_{40k}			
Intercept, γ_{400}	-2.36*	.55	-4.27
School model for classroom level effect of DAP, β_{41k}			
Intercept, γ_{410}	.25	.28	.89
Classroom model for student level effect of HISPANIC, π_{5jk}			
School model for classroom level intercept, β_{50k}			
Intercept, γ_{500}	-1.44*	.47	-3.02
School model for classroom level effect of DAP, β_{51k}			
Intercept, γ_{510}	.26	.23	1.12

Table 23 (continued).

Fixed Effects	*Coefficient*	*se*	*t- ratio*
Classroom model for student level effect of ASIAN, π_{6jk}			
School model for classroom level intercept, β_{60k}			
Intercept, γ_{600}	2.03*	.80	2.53
School model for classroom level effect of DAP, β_{61k}			
Intercept, γ_{610}	.38	.41	.92
Classroom model for student level effect of OTHER, π_{7jk}			
School model for classroom level intercept, β_{70k}			
Intercept, γ_{700}	-.88	.74	-1.19
School model for classroom level effect of DAP, β_{71k}			
Intercept, γ_{710}	.73	.34	2.13

Random Effects	Variance	*df*	χ^2	*p-value*
Schools, u_{00k}	38.61	1214	2916.74	.001
Classrooms, r_{0j}	29.12	2040	4384.18	.001
Students, e_{ijk}	14.65			

*Note: se= standard error of estimate; df=degrees of freedom; χ^2=chi-square statistic; * p<.05.*

In the school model, PUBLIC, indicating school type (sector), was added to the model to investigate the effect of school sector on reading performance. Results of the final (school) model are presented in Table 24. Results of the fixed effects indicated that private schools had higher mean reading achievement than did public schools (γ_{001}= –6.03, p>.05; coded as Public=1, Private=0), but this effect was not statistically significant as indicated. This finding provided evidence that public and private schools did not differ in their reading scores.

There was no evidence of a significant effect of interaction between PUBLIC and DAP on the reading achievement scores (γ_{011}=.30, p>.05). It means that public and private schools did not differ with respect to their average DAP-reading IRT achievement relationship with schools.

Table 24

Results of the Three-level-Analysis of the First Grade IRT Data (School Model)

Fixed Effects	*Coefficient*	*se*	*t- ratio*
Classroom model for student level intercept, π_{0jk}			
School model for classroom level intercept, β_{00k}			
Intercept, γ_{000}	57.02*	.23	246.39
PUBLIC, γ_{001}	-6.03	.57	-1.49
School model for classroom level effect of DAP, β_{01k}			
Intercept, γ_{010}	-.01	.09	-.11
PUBLIC, γ_{011}	.30	.26	1.15
Classroom model for student level effect of MALE, π_{1jk}			
School model for classroom level intercept, β_{10k}			
Intercept, γ_{100}	-2.09*	.28	-7.36
Classroom model for student level effect of AGE, π_{2jk}			
School model for classroom level intercept, β_{20k}			
Intercept, γ_{200}	.16*	.03	4.61
Classroom model for student level effect of SES, π_{3jk}			
School model for classroom level intercept, β_{30k}			
Intercept, γ_{300}	3.59*	.24	14.91
Classroom model for student level effect of BLACK, π_{4jk}			
School model for classroom level intercept, β_{40k}			
Intercept, γ_{400}	-2.37*	.73	-3.23
Classroom model for student level effect of HISPANIC, π_{5jk}			
School model for classroom level intercept, β_{50k}			
Intercept, γ_{500}	-1.47*	.59	-2.48
Classroom model for student level effect of ASIAN, π_{6jk}			
School model for classroom level intercept, β_{60k}			
Intercept, γ_{600}	2.06*	.79	2.61
Classroom model for student level effect of OTHER, π_{7jk}			
School model for classroom level intercept, β_{70k}			
Intercept, γ_{700}	-.87	.88	-.99

Random Effects	Variance	*df*	χ^2	*p-value*
Schools, u_{00k}	33.49	1213	2857.36	.001
Classrooms, r_{0j}	28.97	2040	4535.75	.001
Students, e_{ijk}	14.67			

*Note: se= standard error of estimate; df=degrees of freedom; χ^2=chi-square statistic; * $p<.05$.*

Examination of the variance components placed at the end of the Table 24 indicated that significant variation in the intercepts, including the variance of reading performance between classroom within schools (r_{0jk} =28.97, χ^2=4535.75, *df*=2040, p<.001) and the variance of reading performance between schools (u_{00k} =33.49, χ^2=2857.36, *df*=1213, p<.001) remained unexplained even after controlling for PUBLIC. The proportion of variance explained in y_{000} was estimated as = [38.68-33.49] / 38.68 = .13, indicating that 13% of the parameter variation in mean reading achievement between schools was explained by adding all variables.

First Grade ARS Data

The examination of the fully unconditional model for the first grade ARS data, presented in Table 25, indicated that grand mean of student reading scores was 3.45 (γ_{000}). That is, the weighted least squares estimate for the grand mean reading achievement at the First Grade level was estimated as 3.45 (t=247.09, p<.001) with a standard error of .01 and the 95% confidence interval was;

$$3.45 \pm 1.96\ (.24) = (3.43,\ 3.47).$$

Table 25
Results of the Three-level-Analysis of the First Grade ARS Data (Fully Unconditional Model)

Fixed Effects	*Coefficient*	*se*	*t- ratio*
Classroom model for student level intercept, π_{0jk}			
School model for classroom level intercept, β_{00k}			
Intercept, γ_{000}	3.45*	.01	247.09

Random Effects	Variance	*df*	χ^2	*p-value*
School mean, u_{00k}	.06	1208	1779.33	.001
Classroom mean, r_{0jk}	.20	2075	5954.45	.001
Students, e_{ijk}	.08			

*Note: se= standard error of estimate; df=degrees of freedom; χ^2=chi-square statistic; * p<. 05.*

The chi-square statistics associated with these variance components indicated significant variations among students within classrooms, among classrooms with schools and also between schools. Specifically, the variance components were estimated as .08 (σ^2) among students over classrooms, .20 (τ_π) between classrooms within schools, and .06 (τ_β) between schools. To estimate of the percentage of variance in the reading achievement at each level, the following equations were used;

% variance among students within classrooms: $\sigma^2 / (\sigma^2 + \tau_\pi + \tau_\beta)$,

% variance among classrooms within schools: $\tau_\pi / (\sigma^2 + \tau_\pi + \tau_\beta)$,

% variance between schools: $\tau_\beta / (\sigma^2 + \tau_\pi + \tau_\beta)$,

Computations revealed that 24% of the variance laid between students, 59% of the variance laid between classrooms within schools, and 17% of the total variance laid between schools, respectively. Again, as reported in the Kindergarten ARS data, most of the variance was observed at the classroom level.

The estimate of the between school variability (the variance of school means around the grand-mean) is .06. Therefore, 95% of the school means fall within the following range:

$$3.45 \pm 1.96\,(.06)^{1/2} = (2.97, 3.93)$$

The result of the test for whether all schools have the same mean was found to be significant at .05 level (χ^2=1779.33, p<.001). This means that there is a significant variation between schools in terms of first grade reading achievement in ARS scale.

In the student model, same student level predictors, as in the first grade IRT student model, were added to the first level in order to examine the effects of these variables on students' reading performance and also to explain the variation of reading scores.

The student level model presented in Table 26 displayed that the average school reading ARS achievement mean (grand mean) was 3.45(γ_{000}) (p<.001) for the typical student (Female=0, White=0, average respect to AGE and SES).

Table 26

Results of the Three-level-Analysis of the First Grade ARS Data (Student Model)

Fixed Effects	*Coefficient*	*se*	*t- ratio*
Classroom model for student level intercept, π_{0jk}			
School model for classroom level intercept, β_{00k}			
Intercept, γ_{000}	3.45*	.01	246.70
Classroom model for student level effect of MALE, π_{1jk}			
School model for classroom level intercept, β_{10k}			
Intercept, γ_{100}	-.21*	.01	-12.27
Classroom model for student level effect of AGE, π_{2jk}			
School model for classroom level intercept, β_{20k}			
Intercept, γ_{200}	.02*	.00	8.60
Classroom model for student level effect of SES, π_{3jk}			
School model for classroom level intercept, β_{30k}			
Intercept, γ_{300}	.31*	.01	21.20
Classroom model for student level effect of BLACK, π_{4jk}			
School model for classroom level intercept, β_{40k}			
Intercept, γ_{400}	-.24*	.03	-6.08
Classroom model for student level effect of HISPANIC, π_{5jk}			
School model for classroom level intercept, β_{50k}			
Intercept, γ_{500}	-.17*	.03	-5.19
Classroom model for student level effect of ASIAN, π_{6jk}			
School model for classroom level intercept, β_{60k}			
Intercept, γ_{600}	.04	.05	.79
Classroom model for student level effect of OTHER, π_{7jk}			
School model for classroom level intercept, β_{70k}			
Intercept, γ_{700}	-.09	.05	-1.80

Random Effects	Variance	*df*	χ^2	*p-value*
Schools, u_{00k}	.06	1208	1781.44	.001
Classrooms, r_{0jk}	.22	2075	635.80	.001
Students, e_{ijk}	.07			

*Note: se= standard error of estimate; df=degrees of freedom; χ^2=chi-square statistic; * p<.05.*

Results from the student model's estimated fixed effects indicated that all effects of selected student level predictors, except ASIAN and OTHER, were statistically significant on the reading ARS scores (*p*<.001). In particular, male students' reading achievement on the ARS scale was .21 (γ_{100}) points lower than female students. The relationship between AGE and the reading performance was estimated as .02 (γ_{200}). This finding indicated that a student, who was older one standard deviation, scored .02 points greater than younger student. The effect of SES on the reading IRT achievement was found to be statistically significant (γ_{300} =.31), indicating that a student with one standard deviation higher SES level scored .31 points higher than others. The effect of BLACK on the reading performance was –.24 (γ_{400}), referring that an African-American student scored .24 points lower than his or her counterparts who were Caucasians within school *k*. The effect of HISPANIC on the reading performance was – .17 (γ_{500}), indicating that a Hispanic student scored .17 points lower than Caucasian students as the reference group. The effect of ASIAN on the reading achievement scores was found to be .04 (γ_{600}), indicating that an Asian student scored .04 points higher than his or her counterparts who were Caucasian within school *k*, but this effect was not significant. Finally, the effect of OTHER on the reading performance was found to be –.09 (γ_{700}), indicating that a student who was in the other race group rather than Caucasian, African-American, Hispanic, or Asian, scored .09 points lower than his or her counterparts within school *k*, but again this effect was not significant.

Estimated variances and related chi square statistics from three levels of the student model were presented at the end of the Table 26. These results suggested that even though residual parameter variances in all levels decreased slightly, they still remained unexplained in π_{0jk} (.07), β_{00j} (.22), and γ_{00k} (.06).

In the classroom model, DAP as a classroom level indicator was added to the model to examine the effect of DAP on reading performance and also the effect of the relationship between DAP and the selected student level predictors on reading achievement. Results of the classroom model indicated that DAP had a significant effect on the reading performance (γ_{010}= .01, t= 2.41, p<.05) (See Table 27). On the other hand, the strengths of relations between reading achievement and all of the student level predictors in the model did not differ when considering the effect of DAP.

Table 27

Results of the Three-level-Analysis of the First Grade ARS Data (Classroom Model)

Fixed Effects	*Coefficient*	*se*	*t- ratio*
Classroom model for student level intercept, π_{0jk}			
School model for classroom level intercept, β_{00k}			
Intercept, γ_{000}	3.45*	.01	246.95
School model for classroom level effect of DAP, β_{01k}			
Intercept, γ_{010}	.01*	.00	2.41
Classroom model for student level effect of MALE, π_{1jk}			
School model for classroom level intercept, β_{10k}			
Intercept, γ_{100}	-.21*	.01	-12.38
School model for classroom level effect of DAP, β_{11k}			
Intercept, γ_{110}	-.01	.00	-1.72
Classroom model for student level effect of AGE, π_{2jk}			
School model for classroom level intercept, β_{20k}			
Intercept, γ_{200}	.01*	.00	8.47
School model for classroom level effect of DAP, β_{21k}			
Intercept, γ_{210}	-.00	.00	-.27
Classroom model for student level effect of SES, π_{3jk}			
School model for classroom level intercept, β_{30k}			
Intercept, γ_{300}	.31*	.01	21.22
School model for classroom level effect of DAP, β_{31k}			
Intercept, γ_{310}	.00	.00	.75
Classroom model for student level effect of BLACK, π_{4jk}			
School model for classroom level intercept, β_{40k}			
Intercept, γ_{400}	-.23*	.03	-6.04
School model for classroom level effect of DAP, β_{41k}			
Intercept, γ_{410}	-.01	.02	-.75
Classroom model for student level effect of HISPANIC, π_{5jk}			
School model for classroom level intercept, β_{50k}			
Intercept, γ_{500}	-.17*	.03	-5.12
School model for classroom level effect of DAP, β_{51k}			
Intercept, γ_{510}	.00	.01	.13

Table 27 (continued).

Fixed Effects	*Coefficient*	*se*	*t- ratio*
Classroom model for student level effect of ASIAN, π_{6jk}			
School model for classroom level intercept, β_{60k}			
Intercept, γ_{600}	.04	.05	.76
School model for classroom level effect of DAP, β_{61k}			
Intercept, γ_{610}	.00	.02	.27
Classroom model for student level effect of OTHER, π_{7jk}			
School model for classroom level intercept, β_{70k}			
Intercept, γ_{700}	-.09	.05	-1.77
School model for classroom level effect of DAP, β_{71k}			
Intercept, γ_{710}	.02	.02	.94

Random Effects	Variance	*df*	χ^2	*p-value*
Schools, u_{00k}	.06	1208	1781.54	.001
Classrooms, r_{0j}	.22	2074	6344.88	.001
Students, e_{ijk}	.07			

Note: se= standard error of estimate; df=degrees of freedom; χ^2*=chi-square statistic; * p<.05.*

The estimated variances of the intercepts were slightly smaller than the previous student model, without controlling for DAP. Therefore, significant variation in the intercepts, including the variance of reading performance between classroom within schools (r_{0jk} =.22, χ^2=6418.10, *df*=2074, p<.001) and the variance of reading performance between schools (u_{00k} =.05, χ^2=1729.03, *df*=1207, p<.001) remained unexplained even after controlling for DAP.

In the school model, PUBLIC, indicating school type (sector), was added to the model to examine the effect of school sector on reading performance. Results of the final (school) model are presented in Table 28. Results of the fixed effects indicated that private schools had higher mean reading achievement than did public schools (γ_{001}= –.24, *t*= –5.70, p<.05; coded as Public=1, Private=0).

Table 28

Results of the Three-level-Analysis of the First Grade ARS Data (School Model)

Fixed Effects	*Coefficient*	*se*	*t- ratio*
Classroom model for student level intercept, π_{0jk}			
School model for classroom level intercept, β_{00k}			
Intercept, γ_{000}	3.46*	.01	244.47
PUBLIC, γ_{001}	-.24*	.04	-5.70
School model for classroom level effect of DAP, β_{01k}			
Intercept, γ_{010}	.01*	.00	2.91
PUBLIC, γ_{011}	.03*	.01	2.21
Classroom model for student level effect of MALE, π_{1jk}			
School model for classroom level intercept, β_{10k}			
Intercept, γ_{100}	-.21*	.01	-12.26
Classroom model for student level effect of AGE, π_{2jk}			
School model for classroom level intercept, β_{20k}			
Intercept, γ_{200}	.02*	.00	8.60
Classroom model for student level effect of SES, π_{3jk}			
School model for classroom level intercept, β_{30k}			
Intercept, γ_{300}	.31*	.01	21.20
Classroom model for student level effect of BLACK, π_{4jk}			
School model for classroom level intercept, β_{40k}			
Intercept, γ_{400}	-.24*	.03	-6.08
Classroom model for student level effect of HISPANIC, π_{5jk}			
School model for classroom level intercept, β_{50k}			
Intercept, γ_{500}	-.17*	.03	-5.19
Classroom model for student level effect of ASIAN, π_{6jk}			
School model for classroom level intercept, β_{60k}			
Intercept, γ_{600}	.04	.05	.79
Classroom model for student level effect of OTHER, π_{7jk}			
School model for classroom level intercept, β_{70k}			
Intercept, γ_{700}	-.09	.05	-1.80

Random Effects	Variance	*df*	χ^2	*p-value*
Schools, u_{00k}	.05	1207	1729.03	.001
Classrooms, r_{0j}	.22	2074	6418.10	.001
Students, e_{ijk}	.07			

*Note: se= standard error of estimate; df=degrees of freedom; χ^2=chi-square statistic; * p<.05.*

In addition, there was a significant effect of interaction between PUBLIC and DAP on the reading achievement (γ_{011}=.30, t= 2.21, p<.05). It means that public and private schools differed with respect to their average DAP-reading IRT achievement relationship within schools, favoring public schools over private schools.

Examination of the variance components placed at the end of Table 28 indicated that significant variation in the intercepts, including the variance of reading performance between classroom within schools (r_{0jk} =.22, χ^2=6418.10, df=2074, p<.001) and the variance of reading performance between schools (u_{00k} =.05, χ^2=1729.03, df=1207, p<.001) remained unexplained even after controlling for PUBLIC. The proportion of variance explained in y_{000} was estimated as = [.06-.05] / .06 =.17, indicating that 17% of the parameter variation in mean reading achievement between schools was explained by adding all variables.

Third Grade IRT Data

The results of the fully unconditional model (see Table 29) revealed that grand mean of student reading scores was 111.10 (γ_{000}), indicating that the weighted least squares estimate for the grand mean reading achievement at the third grade level was estimated as 111.10 (t=317.08, p<.001). It had a standard error of .35 and a 95% confidence interval was computed by;

$$95\%\text{CI}(\gamma_{000}) = \gamma \pm 1.96\ (V\gamma_{00})^{1/2}$$

where $V\gamma_{00}$ is the estimated sampling variance of γ_{000}.
The calculation yielded a range of

$$111.10 \pm 1.96\ (.35) = (110.41,\ 111.79).$$

The estimates of the variance components were given in the second part of the Table 32. The chi-square statistics associated with these variance components indicated significant variations among students within classrooms, among classrooms with schools and also between schools. Specifically, the variance components were estimated as 30.59 (σ^2) among students over classrooms, 39.78 (τ_π) between classrooms within schools, and 100.50 (τ_β) between schools.

Table 29
Results of the Three-level-Analysis of the Third Grade IRT Data (Fully Unconditional Model)

Fixed Effects	*Coefficient*	*se*	*t- ratio*
Classroom model for student level intercept, π_{0jk} School model for classroom level intercept, β_{00k} Intercept, γ_{000}	111.10*	.35	317.08

Random Effects	Variance	*df*	χ^2	*p-value*
School mean, u_{00k}	100.50	1489	3736.01	.001
Classroom mean, r_{0jk}	39.78	2096	3394.61	.001
Students, e_{ijk}	30.59			

Note: se= standard error of estimate; df=degrees of freedom; χ^2*=chi-square statistic; * p*<. 05.

The percentage of variance in the reading achievement for each level was computed using the following equations;

% variance among students within classrooms: $\sigma^2 / (\sigma^2 + \tau_\pi + \tau_\beta)$,

% variance among classrooms within schools: $\tau_\pi / (\sigma^2 + \tau_\pi + \tau_\beta)$,

% variance between schools: $\tau_\beta / (\sigma^2 + \tau_\pi + \tau_\beta)$,

Substituting the estimates for each of these variance components into equations, revealed that 18% of the variance laid between students, 24% of the variance laid between classrooms within schools, and 58% of the total variance laid between schools, respectively. Examining the decomposition of variance between levels indicated that most of the variance for the third grade IRT data was observed at the school level.

The estimate of the between school variability (the variance of school means around the grand-mean) is 100.50; therefore, 95% of the school means fall within the following range:

$$111.10 \pm 1.96\,(100.50)^{1/2} = (91.45, 130.75)$$

The result of the test for whether all schools have the same mean was found to be significant at .05 level (χ^2=3736.01, p<.001). This means that there is a significant variation between schools in terms of reading achievement in IRT scale.

Same student level predictors, as in the previous models, were added to the student level in order to investigate the effects of these student characteristics on students' reading performance and also to explain the variation of reading scores. At this grade, because age was a categorical variable in nature, two additional variables were created and replaced with the original AGE variable; they were YOUNGER and OLDER, referring to students who were younger than 9 years and older than 9.5 years old, respectively. The reference group was aged between 9 and 9.5 years of age as the expected chronological age for this grade.

The student level model (see Table 30) indicated that the average school reading IRT achievement mean (grand mean) was 111.08 (γ_{000}) (p<.001) for the typical student (Female=0, White=0, Normal Age=0 [9 years old], average with respect to the SES). The findings regarding the estimated fixed effects indicated that all effects of selected student level predictors were statistically significant on the reading IRT performance (p<.001), except for effects of YOUNGER, ASIAN, and OTHER. Male students' scored 2.96 (γ_{100}) points lower than female students. The relationship between YOUNGER and the reading performance was estimated as .09 (γ_{200}), indicating that a younger student scored .09 points higher than his or her counterparts who were in the norm group within school *k,* yet this effect was not significant. The relationship between OLDER and the reading achievement scores was estimated as 1.57 (γ_{300}), indicating that an older student scored 1.57 points higher than his or her counterparts who were aged between 9 and 9.5 within school *k*. The effect of SES was found to be statistically significant. The estimated effect of SES, (γ_{400}=6.25), indicated that a student with one standard deviation higher SES level scored 6.25 points higher than others. The effect of BLACK on the reading performance was –9.38 (γ_{500}), demonstrating that an African-American student scored 9.38 points lower than his or her Caucasian counterparts within school *k*.

Table 30
Results of the Three-level-Analysis of the Third Grade IRT Data (Student Model)

Fixed Effects	*Coefficient*	*se*	*t- ratio*
Classroom model for student level intercept, π_{0jk}			
School model for classroom level intercept, β_{00k}			
Intercept, γ_{000}	111.08*	.35	316.82
Classroom model for student level effect of MALE, π_{1jk}			
School model for classroom level intercept, β_{10k}			
Intercept, γ_{100}	–2.96*	.43	–6.97
Classroom model for student level effect of YOUNGER, π_{2jk}			
School model for classroom level intercept, β_{20k}			
Intercept, γ_{200}	.09	.53	.18
Classroom model for student level effect of OLDER, π_{3jk}			
School model for classroom level intercept, β_{30k}			
Intercept, γ_{300}	1.57*	.52	2.99
Classroom model for student level effect of SES, π_{4jk}			
School model for classroom level intercept, β_{40k}			
Intercept, γ_{400}	6.25*	.36	17.22
Classroom model for student level effect of BLACK, π_{5jk}			
School model for classroom level intercept, β_{50k}			
Intercept, γ_{500}	–9.38*	1.37	–6.87
Classroom model for student level effect of HISPANIC, π_{6jk}			
School model for classroom level intercept, β_{60k}			
Intercept, γ_{600}	–2.11*	.87	–2.42
Classroom model for student level effect of ASIAN, π_{7jk}			
School model for classroom level intercept, β_{70k}			
Intercept, γ_{700}	–.69	1.48	–.46
Classroom model for student level effect of OTHER, π_{8jk}			
School model for classroom level intercept, β_{80k}			
Intercept, γ_{800}	–.84	1.33	–.63

Random Effects	Variance	*df*	χ^2	*p*-value
Schools, *u*00k	100.49	1489	3711.21	.001
Classrooms, *r*0jk	38.92	1484	3217.61	.001
Students, *e*ijk	28.19			

*Note: se= standard error of estimate; df=degrees of freedom; χ^2=chi-square statistic; * p<.05.*

The effect of HISPANIC was – 2.11 (γ_{600}), referring to a 2.11 points disadvantage against Caucasian students. The effect of ASIAN on the reading performance was found to be -.69 (γ_{700}), indicating that an Asian student scored .69 points lower than his or her counterparts who were in the Caucasian race group within school *k*, however this effect was not significant. Finally, the effect of OTHER on the reading performance was found to be –.84 (γ_{800}), indicating that a student who was in the other race group rather than white, black, Hispanic, or Asian, scored .84 points lower than his or her counterparts within school *k,* but this effect was not significant.

Estimated variances and related chi square statistics from three level of the student model were presented at the end of the Table 30. These results suggested that even though residual parameter variances in all levels decreased slightly, they still remained unexplained in π_{0jk} (28.19), β_{00j} (39.92), and γ_{00k} (100.49).

Inclusion of DAP in the classroom model permitted the examination of the effect of DAP on reading performance as well as the effect of the relationship between DAP and the selected student level predictors. Results of the classroom model indicated that DAP did not have a significant effect on the reading performance (γ_{010}=111.07, t= 316.76, p<.01) (See Table 31) assessed by the standardized direct measure. The strengths of associations between reading achievement and most of the student level predictors in the model did not differ when considering the effect of DAP, except for the interaction effect between BLACK and DAP on performance. In other words, most of the interaction effects between selected student characteristics and DAP were not found to be significant on the IRT reading achievement scores at .05 level. However, the effect of DAP on the effect of BLACK was statistically significant and positive in direction (γ_{510}= 1.41, t= 2.03, p< .05). It means that African-American students benefited more from DAP implications in terms of reading achievement. African-American students experienced a 1.41-point advantage over Caucasian students in higher DAP classrooms.

Table 31
Results of the Three-level-Analysis of the Third Grade IRT Data (Classroom Model)

Fixed Effects	*Coefficient*	*se*	*t- ratio*
Classroom model for student level intercept, π_{0jk}			
School model for classroom level intercept, β_{00k}			
Intercept, γ_{000}	111.07*	.35	316.76
School model for classroom level effect of DAP, β_{01k}			
Intercept, γ_{010}	.08	.18	.46
Classroom model for student level effect of MALE, π_{1jk}			
School model for classroom level intercept, β_{10k}			
Intercept, γ_{100}	-2.93*	.42	-6.88
School model for classroom level effect of DAP, β_{11k}			
Intercept, γ_{110}	-.15	.30	-.51
Classroom model for student level effect of YOUNGER, π_{2jk}			
School model for classroom level intercept, β_{20k}			
Intercept, γ_{200}	.24	.52	.45
School model for classroom level effect of DAP, β_{21k}			
Intercept, γ_{210}	-.26	.37	-.70
Classroom model for student level effect of OLDER, π_{3jk}			
School model for classroom level intercept, β_{30k}			
Intercept, γ_{300}	1.66*	.52	3.16
School model for classroom level effect of DAP, β_{31k}			
Intercept, γ_{310}	.38	.37	1.03
Classroom model for student level effect of SES, π_{4jk}			
School model for classroom level intercept, β_{40k}			
Intercept, γ_{400}	6.21*	.36	17.14
School model for classroom level effect of DAP, β_{41k}			
Intercept, γ_{410}	.25	.25	1.02
Classroom model for student level effect of BLACK, π_{5jk}			
School model for classroom level intercept, β_{50k}			
Intercept, γ_{500}	-9.15*	1.10	-8.31
School model for classroom level effect of DAP, β_{51k}			
Intercept, γ_{510}	1.41*	.69	2.03

Table 31 (continued).

Fixed Effects	*Coefficient*	*se*	*t- ratio*
Classroom model for student level effect of HISPANIC, π_{6jk}			
School model for classroom level intercept, β_{60k}			
Intercept, γ_{600}	-1.96*	.87	-2.25
School model for classroom level effect of DAP, β_{61k}			
Intercept, γ_{610}	-.29	.59	-.50
Classroom model for student level effect of ASIAN, π_{7jk}			
School model for classroom level intercept, β_{70k}			
Intercept, γ_{700}	-.58	1.49	-.39
School model for classroom level effect of DAP, β_{71k}			
Intercept, γ_{710}	-1.28	1.11	-1.14
Classroom model for student level effect of OTHER, π_{8jk}			
School model for classroom level intercept, β_{80k}			
Intercept, γ_{800}	-.91	1.33	-.68
School model for classroom level effect of DAP, β_{71k}			
Intercept, γ_{810}	-1.07	.90	-1.19

Random Effects	Variance	*df*	χ^2	*p-value*
Schools, u_{00k}	101.32	1489	3754.15	.001
Classrooms, r_{0j}	37.12	1483	3124.54	.001
Students, e_{ijk}	28.19			

Note: se= standard error of estimate; df=degrees of freedom; χ^2*=chi-square statistic; * p<.05.*

The estimated variances of the intercepts were slightly smaller than the previous student model, without controlling for DAP. Therefore, significant variation in the intercepts, including the variance of reading performance between classroom within schools (r_{0jk} =37.12, χ^2=3124.54, *df*=2096, p<.001) and the variance of reading performance between schools (u_{00k} =.05, χ^2=3754.15, *df*=1489, p<.001) remained unexplained even after controlling for DAP.

In the school model, PUBLIC, indicating school type (sector) was added to the model to examine the effect of school sector on reading performance. Results of the fixed effects (see Table 32) in the final model indicated that private schools had higher mean reading achievement than did public schools (γ_{001}=–7.25, t=–8.88, *p*<.01; coded as Public=1, Private=0 in the data).

Table 32

Results of the Three-level-Analysis of the Third Grade IRT Data (School Model)

Fixed Effects	*Coefficient*	*se*	*t- ratio*
Classroom model for student level intercept, π_{0jk}			
School model for classroom level intercept, β_{00k}			
Intercept, γ_{000}	111.24*	.34	324.98
PUBLIC, γ_{001}	-7.25*	.81	-8.88
School model for classroom level effect of DAP, β_{01k}			
Intercept, γ_{010}	.17	.17	1.00
PUBLIC, γ_{011}	-.39	.42	-.93
Classroom model for student level effect of MALE, π_{1jk}			
School model for classroom level intercept, β_{10k}			
Intercept, γ_{100}	-2.91*	.51	-5.70
Classroom model for student level effect of YOUNGER, π_{2jk}			
School model for classroom level intercept, β_{20k}			
Intercept, γ_{200}	.21	.60	.34
Classroom model for student level effect of OLDER, π_{3jk}			
School model for classroom level intercept, β_{30k}			
Intercept, γ_{300}	1.59*	.68	2.32
Classroom model for student level effect of SES, π_{4jk}			
School model for classroom level intercept, β_{40k}			
Intercept, γ_{400}	6.25*	.43	14.24
Classroom model for student level effect of BLACK, π_{5jk}			
School model for classroom level intercept, β_{50k}			
Intercept, γ_{500}	-9.16*	1.39	-6.58
School model for classroom level effect of DAP, β_{51k}			
Intercept, γ_{510}	1.53*	.72	2.12
PUBLIC, γ_{511}	.14	2.47	.05
Classroom model for student level effect of ASIAN, π_{7jk}			
School model for classroom level intercept, β_{70k}			
Intercept, γ_{700}	-2.01	1.07	-1.86
Classroom model for student level effect of OTHER, π_{8jk}			
School model for classroom level intercept, β_{80k}			
Intercept, γ_{800}	-.69	1.29	-.54

Random Effects	Variance	*df*	χ^2	*p*-value
Schools, u_{00k}	94.13	1488	3611.39	.001
Classrooms, r_{0jk}	39.10	1483	3317.78	.001
Students, e_{ijk}	28.18			

*Note: se= standard error of estimate; df=degrees of freedom; χ^2=chi-square statistic; * p<.05.*

There was no evidence of a significant effect of interaction between PUBLIC and DAP on the reading achievement scores (γ_{011}=-.39, p>.05). It means that public and private schools did not differ with respect to their average DAP-reading IRT achievement relationship with schools. Similarly, the effect of PUBLIC on the effect of BLACK-reading achievement slope (γ_{511}=.14, p>.05) was not statistically significant at .05 level, indicating that public and private schools did not vary with respect to their average DAP effect on BLACK-reading IRT achievement relationship with schools.

Examination of the variance components indicated that significant variation in the intercepts, including the variance of reading performance between classroom within schools (r_{0jk} =39.10, χ^2=3317.78, *df*=2096, p<.001) and the variance of reading performance between schools (u_{00k} =94.13, χ^2=3611.39, *df*=1488, p<.001) remained unexplained even after controlling for PUBLIC. The proportion of variance explained in y_{000} was estimated as = [100.50 – 94.13] / 100.50=.07, indicating that 7% of the parameter variation in mean reading achievement between schools was explained by adding all variables.

Third Grade ARS Data

The results of the fully unconditional model, presented in Table 33, showed that grand mean of student reading scores was 3.42 (γ_{000}). In other words, the weighted least squares estimate for the grand mean reading achievement at the third grade level was estimated as 111.10 (*t*=248.73, *p*<.001). It had a standard error of .01 and a 95% confidence interval was computed by;

$$95\%CI(\gamma_{000}) = \gamma \pm 1.96\ (V\gamma_{00})^{1/2}$$

where $V\gamma_{00}$ is the estimated sampling variance of γ_{000}.

The calculation yielded a range of

$$3.42 \pm 1.96\ (.01) = (3.40, 3.44).$$

Table 33
Results of the Three-level-Analysis of the Third Grade ARS Data (Fully Unconditional Model)

Fixed Effects	*Coefficient*	*se*	*t*- ratio
Classroom model for student level intercept, π_{0jk}			
School model for classroom level intercept, β_{00k}			
Intercept, γ_{000}	3.42*	.01	245.73

Random Effects	Variance	*df*	χ^2	*p*-value
School mean, u_{00k}	.04	1463	1781.27	.001
Classroom mean, r_{0jk}	.25	1435	6400.21	.001
Students, e_{ijk}	.06			

*Note: se= standard error of estimate; df=degrees of freedom; χ^2=chi-square statistic; * $p<.05$.*

The estimates of the variance components were given in the second part of the Table 33. The chi-square statistics associated with these variance components indicated significant variations among students within classrooms, among classrooms with schools and also between schools. Specifically, the variance components were estimated as .06 (σ^2) among students over classrooms, .25 (τ_π) between classrooms within schools, and .04 (τ_β) between schools. To estimate of the percentage of variance in the reading achievement at each level, the following equations were used;

% variance among students within classrooms: $\sigma^2 / (\sigma^2 + \tau_\pi + \tau_\beta)$,

% variance among classrooms within schools: $\tau_\pi / (\sigma^2 + \tau_\pi + \tau_\beta)$,

% variance between schools: $\tau_\beta / (\sigma^2 + \tau_\pi + \tau_\beta)$,

Substituting the estimates for each of these variance components into equations, revealed that 17% of the variance laid between students, 71% of the variance laid between classrooms within schools, and 12% of the total variance laid between schools, respectively. Examining the decomposition of variance between levels indicated that most of the variance for the third grade ARS data was observed at the classroom level. The estimate of the between school variability (the variance of school means around the grand-mean) is 3.42; therefore, 95% of the school means fall within the following range:

$$3.42 \pm 1.96\,(.04)^{1/2} = (3.03, 3.81)$$

The result of the test for whether all schools have the same mean was found to be significant at .05 level (2=1781.27, p<.001). The findings demonstrated a significant variation between schools in reading achievement by the ARS scale.

In the student model, same student level predictors, as in the previous model of the Third grade IRT data, were added to the student level in order to investigate the effects of these variables on students' reading performance and also to explain the variation of reading scores.

The findings regarding the student level model presented in Table 34 exhibited that the average school reading ARS achievement mean (grand mean) was 3.42 (γ_{000}) (p<.001) for the typical student (Female=0, White=0, Normal Age=0 [9 years old], average with respect to the SES). Results from the student model's estimated fixed effects indicated that some of the effects of selected student level predictors were statistically significant on the reading ARS performance (p<.001), except for effects of MALE, OLDER, ASIAN, and OTHER. In an examination of the effects of gender, male students' reading achievement on the ARS scale was .20 (γ_{100}) points lower than female students, yet this effect was not significant. The relationship between YOUNGER and the reading scores was estimated as .30 (γ_{200}), indicating that a younger student scored .30 points higher than his or her counterparts who were in the other age groups within school *k.* The relationship between OLDER and the reading performance was estimated as –.33 (γ_{300}), indicating that an older student scored .33 points lower than his or her counterparts who were in the other age groups within school *k,* but this effect was also not significant. The effect of SES on the estimated reading IRT achievement was found to be statistically significant. The estimated effect of SES, (γ_{400}=.11), indicated that a student with one standard deviation higher SES level scored .11 points higher than others.

Table 34

Results of the Three-level-Analysis of the Third Grade ARS Data (Student Model)

Fixed Effects	*Coefficient*	*se*	*t- ratio*
Classroom model for student level intercept, π_{0jk}			
School model for classroom level intercept, β_{00k}			
Intercept, γ_{000}	3.42*	.01	246.58
Classroom model for student level effect of MALE, π_{1jk}			
School model for classroom level intercept, β_{10k}			
Intercept, γ_{100}	-.20	.01	-1.54
Classroom model for student level effect of YOUNGER, π_{2jk}			
School model for classroom level intercept, β_{20k}			
Intercept, γ_{200}	.30*	.01	18.01
Classroom model for student level effect of OLDER, π_{3jk}			
School model for classroom level intercept, β_{30k}			
Intercept, γ_{300}	-.03	.02	-1.29
Classroom model for student level effect of SES, π_{4jk}			
School model for classroom level intercept, β_{40k}			
Intercept, γ_{400}	.11*	.02	4.52
Classroom model for student level effect of BLACK, π_{5jk}			
School model for classroom level intercept, β_{50k}			
Intercept, γ_{500}	-.37*	.05	-7.13
Classroom model for student level effect of HISPANIC, π_{6jk}			
School model for classroom level intercept, β_{60k}			
Intercept, γ_{600}	-.11*	.04	-2.81
Classroom model for student level effect of ASIAN, π_{7jk}			
School model for classroom level intercept, β_{70k}			
Intercept, γ_{700}	.12	.07	1.73
Classroom model for student level effect of OTHER, π_{8jk}			
School model for classroom level intercept, β_{80k}			
Intercept, γ_{800}	.04	.06	.77

Random Effects	Variance	*df*	χ^2	*p-value*
Schools, u_{00k}	.03	1463	1756.96	.001
Classrooms, r_{0jk}	.26	1435	6981.26	.001
Students, e_{ijk}	.05			

*Note: se= standard error of estimate; df=degrees of freedom; χ^2=chi-square statistic; * p<.05.*

The effect of BLACK on the reading achievement was –.37 (γ_{500}), indicating that an African-American student scored .37 points lower than his or her counterparts who were Caucasian, the reference group, within school *k*. The effect of HISPANIC was – .11 (γ_{600}), referring to a .11-points disadvantage for a Hispanic student against Caucasian students. The effect of ASIAN on the ARS reading achievement scores was nonsignificant and found to be .12 (γ_{700}), indicating that an Asian student scored .12 points higher than his or her Caucasian counterparts who were in the other race groups within school *k*. Finally, the effect of OTHER on the reading performance was found to be .04 (γ_{800}), indicating that a student who was in the other race group rather than white, black, Hispanic, or Asian, scored .04 points lower than his or her counterparts within school *k,* but this effect was not significant.

Estimated variances and related chi square statistics from three levels of the student model were also presented in Table 34. These results revealed that even though residual parameter variances in all levels decreased slightly, they still remained unexplained in π_{0jk} (.05), β_{00j} (.26), and γ_{00k} (.03).

DAP as a classroom level predictor was added to the classroom model to investigate the effect of DAP on reading performance and also the effect of the relationship between DAP and the selected student level predictors on reading achievement. The findings indicated that DAP had a significant effect on the reading performance (γ_{010}=.04, t= 4.82, p<.01) (See Table 35). The strengths of associations between reading achievement and most of the student level predictors in the model did not differ as a function of DAP, except for the interaction effect between HISPANIC and DAP on performance. In other words, DAP did not exhibit differential effects across the student characteristics included in the model. However, the effect of DAP on the effect of HISPANIC was statistically significant and negative in direction (γ_{510}= –.06, t= –2.21, p< .05), indicating that Hispanic students lower than Caucasians in higher DAP classrooms in terms of reading achievement.

Table 35

Results of the Three-level-Analysis of the Third Grade ARS Data (Classroom Model)

Fixed Effects	*Coefficient*	*se*	*t- ratio*
Classroom model for student level intercept, π_{0jk}			
School model for classroom level intercept, β_{00k}			
Intercept, γ_{000}	3.42	.01	248.10
School model for classroom level effect of DAP, β_{01k}			
Intercept, γ_{010}	.04	.00	4.82
Classroom model for student level effect of MALE, π_{1jk}			
School model for classroom level intercept, β_{10k}			
Intercept, γ_{100}	-.20	.01	-1.51
School model for classroom level effect of DAP, β_{11k}			
Intercept, γ_{110}	.00	.01	.42
Classroom model for student level effect of YOUNGER, π_{2jk}			
School model for classroom level intercept, β_{20k}			
Intercept, γ_{200}	.30	.01	17.94
School model for classroom level effect of DAP, β_{21k}			
Intercept, γ_{210}	-.00	.01	-.00
Classroom model for student level effect of OLDER, π_{3jk}			
School model for classroom level intercept, β_{30k}			
Intercept, γ_{300}	-.02	.02	-1.18
School model for classroom level effect of DAP, β_{31k}			
Intercept, γ_{310}	.00	.01	.25
Classroom model for student level effect of SES, π_{4jk}			
School model for classroom level intercept, β_{40k}			
Intercept, γ_{400}	.11	.02	4.73
School model for classroom level effect of DAP, β_{41k}			
Intercept, γ_{410}	.03	.01	1.85
Classroom model for student level effect of BLACK, π_{5jk}			
School model for classroom level intercept, β_{50k}			
Intercept, γ_{500}	-.37	.05	-7.18
School model for classroom level effect of DAP, β_{51k}			
Intercept, γ_{510}	.00	.03	.04

Table 35 (continued).

Fixed Effects	*Coefficient*	*se*	*t- ratio*
Classroom model for student level effect of HISPANIC, π_{6jk}			
School model for classroom level intercept, β_{60k}			
Intercept, γ_{600}	-.11	.04	-2.86
School model for classroom level effect of DAP, β_{61k}			
Intercept, γ_{610}	-.06	.02	-2.21
Classroom model for student level effect of ASIAN, π_{7jk}			
School model for classroom level intercept, β_{70k}			
Intercept, γ_{700}	.12	.07	1.75
School model for classroom level effect of DAP, β_{71k}			
Intercept, γ_{710}	-.05	.05	-1.11
Classroom model for student level effect of OTHER, π_{8jk}			
School model for classroom level intercept, β_{80k}			
Intercept, γ_{800}	.05	.06	.84
School model for classroom level effect of DAP, β_{71k}			
Intercept, γ_{810}	-.08	.04	-1.91

Random Effects	Variance	*df*	χ^2	*p-value*
Schools, u_{00k}	.03	1463	1751.85	.001
Classrooms, r_{0jk}	.26	1434	6973.24	.001
Students, e_{ijk}	.05			

Note: se= standard error of estimate; df=degrees of freedom; χ^2*=chi-square statistic; ** $p<.05$.

The estimated variances of the intercepts were slightly smaller than the previous student model, without controlling for DAP. Therefore, significant variation in the intercepts, including the variance of reading performance between classroom within schools (r_{0jk} =.26, χ^2=6973.24, *df*=2078, p<.001) and the variance of reading performance between schools (u_{00k} =.03, χ^2=1751.85, *df*=1460, p<.001) remained unexplained even after controlling for DAP.

In the school model, PUBLIC, indicating school type (sector), was added to the model to examine the effect of school sector on reading performance. Results of the final (school) model are presented in Table 36.

Table 36

Results of the Three-level-Analysis of the Third Grade ARS Data (School Model)

Fixed Effects	*Coefficient*	*se*	*t- ratio*
Classroom model for student level intercept, π_{0jk}			
School model for classroom level intercept, β_{00k}			
Intercept, γ_{000}	3.42*	.01	245.38
PUBLIC, γ_{001}	-.06	.04	-1.69
School model for classroom level effect of DAP, β_{01k}			
Intercept, γ_{010}	.04*	.00	4.93
PUBLIC, γ_{011}	-.02	.02	-1.03
Classroom model for student level effect of MALE, π_{1jk}			
School model for classroom level intercept, β_{10k}			
Intercept, γ_{100}	-.20	.01	-1.54
Classroom model for student level effect of YOUNGER, π_{2jk}			
School model for classroom level intercept, β_{20k}			
Intercept, γ_{200}	.30*	.01	18.01
Classroom model for student level effect of OLDER, π_{3jk}			
School model for classroom level intercept, β_{30k}			
Intercept, γ_{300}	-.03	.02	-1.29
Classroom model for student level effect of SES, π_{4jk}			
School model for classroom level intercept, β_{40k}			
Intercept, γ_{400}	.11*	.02	4.52
Classroom model for student level effect of BLACK, π_{5jk}			
School model for classroom level intercept, β_{50k}			
Intercept, γ_{500}	-.37*	.05	-7.13
Classroom model for student level effect of HISPANIC, π_{6jk}			
School model for classroom level intercept, β_{60k}			
Intercept, γ_{600}	-.11*	.04	-2.81
School model for classroom level effect of DAP, β_{6k}			
Intercept, γ_{601}	-.05	.02	-1.85
PUBLIC, γ_{611}	.04	.06	.60
Classroom model for student level effect of ASIAN, π_{7jk}			
School model for classroom level intercept, β_{70k}			
Intercept, γ_{700}	.12	.07	1.72
Classroom model for student level effect of OTHER, π_{8jk}			
School model for classroom level intercept, β_{80k}			
Intercept, γ_{800}	.04	.06	.77

Random Effects	Variance	*df*	χ^2	*p-value*
Schools, u_{00k}	.03	1462	1742.55	.001
Classrooms, r_{0j}	.26	1434	6968.92	.001
Students, e_{ijk}	.05			

*Note: se= standard error of estimate; df=degrees of freedom; χ^2=chi-square statistic; * p<.05.*

Results of the fixed effects indicated that private schools had higher mean reading achievement than did public schools (γ_{001}=–.06, t= –1.69, p>.05; coded as Public=1, Private=0 in the data). Also, there was no evidence of a significant effect of interaction between PUBLIC and DAP on the ARS reading achievement scores (γ_{011}= –.02, t= –1.03 p>.05). It means that public and private schools did not differ with respect to their average DAP-reading ARS achievement relationship within schools. Similarly, the effect of PUBLIC on the effect of HISPANIC-reading achievement slope (γ_{611}=.04, t= .60, p >.05) was not statistically significant at .05 level, indicating that public and private schools did not vary with respect to their average DAP effect on HISPANIC-reading ARS achievement relationship within schools.

Examination of the variance components indicated that significant variation in the intercepts, including the variance of reading performance between classroom within schools (r_{0jk} =.26, χ^2=6968.92, df=2078, p<.001) and the variance of reading performance between schools (u_{00k} =.03, χ^2=1742.55, df=1459, p<.001) remained unexplained even after controlling for PUBLIC. The proportion of variance explained in y_{000} was estimated as = [.04-.03] / .04=.25, indicating that 25% of the parameter variation in mean reading achievement between schools was explained by adding all variables.

Results of Three-Level Hierarchical Linear Growth Modeling

Kindergarten IRT Data

Fully Unconditional Model (Null Model)

Level -1 Model: At level-1 with an individual growth rate of reading achievement at time *t* of student *i* in classroom *j*:

$$\text{Reading (IRT)}_{tjj} = \pi_{0ij} + \pi_{1ij}(\text{TIME})_{tij} + e_{tij}$$

Where;

π_{0ij} : Spring (at Kindergarten) reading achievement for i^{th} individual in classroom j

π_{1ij} : Student growth rate of reading achievement for i^{th} individual in classroom j

e_{tij} : Error variance

Unconditional Level 2 (Student level)

$\pi_{0ij} = \beta_{00j} + r_{0ij}$

$\pi_{1ij} = \beta_{10j} + r_{1ij}$

Unconditional Level 3 (Classroom Level)

$\beta_{00j} = \gamma_{000} + u_{00j}$

$\beta_{10j} = \gamma_{100} + u_{10j}$

The results of the fully unconditional model given in Table 37, did not indicate strong overall growth trajectory averaged across all children and classrooms. The estimated mean intercept, γ_{000}, was 27.34. The average reading learning rate within two semesters from Fall to Spring was estimated at .03 (γ_{100}), indicating that the average reading performance rate was statistically different from zero ($p<.001$).

Table 37

Three-Level Analysis of Kindergarten IRT Reading Data (Fully Unconditional Model)

Fixed Effect	Coefficient	se	*t* ratio
Average final status, γ_{000}	27.34	.12	222.55
Average growth rate, γ_{100}	.03	.00	88.88

Random Effect	Variance	*df*	χ^2	*p*-value
Level 1				
Temporal variation, e_{tij}	12.95			
Level 2 (students within classrooms)				
Individual Final status r_{0ij}	29.25	13914	21918.85	.001
Individual Growth rate, r_{1ij}	.00	13914	3293.30	>.500
Level 3 (between classrooms)				
Classroom mean status, u_{ooj}	7.41	967	2296.19	.001
Classroom mean growth rate, u_{1oj}	.00	967	479.57	>.500

Level 1 Coefficient	Percentage of Variance Between Classrooms
Final status, π_{0i}	.20
Growth rate, π_{1i}	–

Variance Covariance Components and Correlations among the Level-2 and Level-3 Random Effects

Level 2 $\begin{pmatrix} 29.25 & .98 \\ .02 & .00 \end{pmatrix} = \hat{\mathrm{T}}\pi = \begin{pmatrix} \hat{\tau}_{\pi 11} & \\ \hat{\tau}_{\pi 12} & \hat{\tau}_{\pi 22} \end{pmatrix}$

Level 3 $\begin{pmatrix} 7.41 & .92 \\ .01 & .00 \end{pmatrix} = \hat{\mathrm{T}}\beta = \begin{pmatrix} \hat{\tau}_{\beta 11} & \\ \hat{\tau}_{\beta 12} & \hat{\tau}_{\beta 22} \end{pmatrix}$

Note: the lower triangles contain the covariances; the upper triangles contain the correlations

**p<.05*

Of substantive interest in this study was the decomposition of the variance in π_{0ij} and π_{1ij} into their within and between classroom components. The estimates for these variance components are presented in the second part of the Table 37. The chi-squared statistics associated with these variance components indicated a significant variation among students within classrooms for final status (π_{0ij}) and a significant variation between classrooms for mean final status (β_{00j}). However, there was no evidence that reading growth rate (π_{1ij}) varied among students within classrooms for (p>.500) and also mean reading learning growth rates (β_{10j}) varied between classrooms (p>.500). That is, hypotheses $\hat{\tau}_{\pi 22} = \hat{V}ar(r_{1ij}) = 0$ and $\hat{\tau}_{\beta 22} = \hat{V}ar(u_{1ij}) = 0$ were sustained.

Based on these variance component estimates, the percentage of variation was computed as falling between classrooms for both final status and growth rates.

$$\text{\% Variance between classrooms on } \pi_{pjk} = \frac{\tau_{\beta pp}}{\tau_{\beta pp} + \tau_{\pi pp}}$$

Substituting corresponding estimates for the variance components into above equation yields the results [as following] presented in the third part of the Table 37;

% Variance between classrooms on π_{0ij} = (7.41) / (7.41+29.25)= .20

% Variance between classrooms on π_{1ij} = undefined

Therefore, 20 % of the variance in final status lies between classrooms while no variance was observed between classrooms for the growth rates. The correlation between final status and growth rate into its within (level 2) and between-classroom (level 3) components were also reported in Table 40 (See the last panel of the table). Within-and between classrooms, both estimated correlations were quite high in magnitude as .98 and .92, respectively, and both values referred positive relations at the within-and between-classroom levels.

In general, the variance component decomposition indicated an important feature of the kindergarten IRT data that a high proportion of the variation in reading learning rates laid within classrooms. The reliability estimates also suggest determining whether certain random effects in within-and-between classrooms models might be constrained to zero. Estimated reliabilities for π_{0ij} (.38) and β_{00j} (.51) were evidently higher than for the π_{1ij} (.01) and β_{10j} (.04). Therefore, random effects of growth rates in reading among students and classrooms were fixed to be zero in the next models.

Student Model (Conditional at Level 2 and Unconditional at Level 3)

In the first conditional model, student level variables were added to model that allowed estimation of the effects of these student characteristics on individual reading achievement. The level 1 model remained same.

Level 1(Student Level)

$$\text{Reading (IRT)}_{tjj} = \pi_{0ij} + \pi_{1ij}(\text{TIME})_{tij} + e_{tij}$$

The level-2 model represents the variability in final status among students within classrooms. Therefore, the random effect of growth rates among students was fixed in the second level based on the results of the fully unconditional model, indicating that there was no variation among students within classrooms in terms of reading growth rate. The level–2 model was formulated as follows. There were only one random effects within classroom (r_{0ij}).

Conditional Level 2 (Student Level)

$$\pi_{0ij} = \beta_{00j} + \beta_{01j}(\text{MALE}) + \beta_{02j}(\text{AGE}) + \beta_{03j}(\text{SES}) + \beta_{04j}(\text{BLACK}) +$$
$$\beta_{05j}(\text{HISPANIC}) + \beta_{06j}(\text{ASIAN}) + \beta_{07j}(\text{OTHER})\ r_{0ij}$$

$$\pi_{1ij} = \beta_{10j}$$

where;

β_{00j} : Estimated mean reading achievement at the end of program within classroom j (final typical student achievement)

β_{01j} : Within-classroom effect of MALE on the estimated mean reading achievement at the end of program for an individual

β_{02j} : Within-classroom effect of AGE on the estimated mean reading achievement at the end of program for an individual

β_{03j} : Within-classroom effect of SES on the estimated mean reading achievement at the end of program for an individual

β_{04j} : Within-classroom effect of BLACK on the estimated mean reading achievement at the end of program for an individual

β_{05j} : Within-classroom effect of Gender HISPANIC on the estimated mean reading achievement at the end of program for an individual

β_{06j} : Within-classroom effect of ASIAN on the estimated mean reading achievement at the end of program for an individual

β_{07j} : Within-classroom effect of OTHER on the estimated mean reading achievement at the end of program for an individual

r_{0ij} :Variance of reading final scores within classrooms

However, the level-3 was still unconditional, meaning that it did not include any classroom level predictors at this time. Random effects of mean reading performance and the mean SES between classrooms were freely estimated but, remaining random effects associated with average student characteristics were fixed to zero based on results of the preliminary analysis. Now the Level-3 was modeled;

<u>Unconditional Level 3 (Classroom Level)</u>

$\beta_{00j} = \gamma_{000} + u_{00j}$

$\beta_{01j} = \gamma_{010}$

$\beta_{02j} = \gamma_{020}$

$\beta_{03j} = \gamma_{020} + u_{02j}$

$\beta_{04j} = \gamma_{040}$

$\beta_{05j} = \gamma_{050}$

$\beta_{06j} = \gamma_{060}$

$\beta_{07j} = \gamma_{070}$

$\beta_{10j} = \gamma_{100j}$

Results from the student model's estimated fixed effects, presented in Table 38, revealed that there were significant relationships between all selected students characteristics and the final reading status. Estimated mean reading achievement at the end of program for an individual who was average with respect to student predictors within a classroom was 27.6. The γ_{000} coefficient represented in this study predicted the final status for a typical child (Female=0, White=0, average in AGE and SES). For such a student, the predicted reading achievement at the kindergarten level was 27.6. Overall mean growth rate of reading achievement during the program was still low, .03. In other words, from fall to spring semesters at the kindergarten level, children's reading achievement increased .031 points.

Table 38
Effects of Student-Level Characteristics on Student Kindergarten Reading Performance (IRT) (Conditional Level-2 Model)

Fixed Effect	Coefficient	se	*t* ratio
Model for final status, π_{0ij}			
Model for mean status of a typical child, β_{00j}			
Intercept γ_{000}	27.60*	.10	287.57
Model for MALE on final status β_{01j} intercept γ_{010}	-1.69*	.10	-13.14
Model for AGE on final status β_{02j} intercept γ_{020}	.37*	.02	23.02
Model for SES on final status β_{03j} intercept γ_{030}	4.29*	.12	36.75
Model for BLACK on final status β_{04j} intercept γ_{040}	-1.44*	.20	-7.39
Model for HISPANIC on final status β_{05j} intercept γ_{050}	-1.28*	.21	-6.26
Model for ASIAN on final status β_{06j} intercept γ_{060}	2.20*	.32	6.95
Model for OTHER on final status β_{07j} intercept γ_{070}	-1.81*	.32	-5.65
Model for growth rate, π_{1ij}			
Intercept γ_{10j} Intercept γ_{100}	.03*	.00	194.81

**p*<.05

Within-classroom effect of MALE on the reading achievement at the end of program for an individual was found to be -1.69 (γ_{010}), indicating that male students ended the program -1.69 points behind their female counterparts (classroom mates), after controlling for all other student characteristics. The effect of AGE on the reading performance at the end of program for an individual was found to be .37 (γ_{020}), indicating that a student, who is older one standard deviation, ended the program .37 points ahead of younger students within classroom j. The effect of SES on the reading scores at the end of program for an individual was found to be 4.29 (γ_{030}), that is a student with one standard deviation higher SES level ended the program 4.29 points higher than others. The effect of BLACK on the IRT reading achievement scores at the end of program for an individual was found to be -1.44 (γ_{040}), referring to a 1.44-points disadvantage for an African-American student against Caucasian children within classroom *j*. The effect of HISPANIC on the estimated mean reading performance at the end of program for an individual was found to be -1.28 (γ_{050}), indicating that an Hispanic student ended the program 1.28 points behind his or her Caucasian counterparts within classroom *j*. On the other hand, the effect of ASIAN at the end of program for an individual was found to be 2.20 (γ_{060}), indicating that an Asian student ended the program 2.20 points ahead of students who were Caucasian within classroom *j*. Finally, the effect of OTHER on the reading performance at the end of program for an individual was found to be -1.81 (γ_{070}), indicating that a student who was in the other race group rather than white, black, Hispanic, or Asian, ended the program 1.81 points behind his or her Caucasian peers within classroom *j*.

Estimated variances and related chi squared statistics from the three level decomposition are illustrated in Table 39 illustrates. These results suggest that residual parameter variances still remained unexplained in π_{0ij}, β_{00j}, and β_{02j}.

Table 39
Variance Decomposition from a three level analysis of the effects of selected student level variables on student reading learning

Random Effect	Variance	*df*	χ^2	*p*-value
Level 1				
Temporal variation, e_{tij}	2.86			
Level 2 (students within classrooms)				
Individual Final status r_{0ij}	47.30	12940	59769.01	.001
Level 3 (between classrooms)				
Classroom mean status, u_{00j}	3.48	945	1667.33	.001
Classroom mean status/SES u_{03j}	3.31	945	137.44	.001

Classroom model (Fully conditional model)

For this model, a classroom level continuous variable "DAP" was added to model to examine the effect of DAP on students' final reading performance and the effect of the relationship between SES and DAP on students' final reading performance.

Conditional Level 2

$\pi_{0ij} = \beta_{00j} + \beta_{01j}\,(\text{MALE}) + \beta_{02j}\,(\text{AGE}) + \beta_{03j}\,(\text{SES}) + \beta_{04j}\,(\text{BLACK}) +$

$\quad \beta_{05j}\,(\text{HISPANIC}) + \beta_{06j}\,(\text{ASIAN}) + \beta_{07j}\,(\text{OTHER})\; r_{0ij}$

$\pi_{1ij} = \beta_{10j}$

Conditional Level 3

$\beta_{00j} = \gamma_{000} + \gamma_{001}\,(\text{DAP}) + u_{00j}$

$\beta_{01j} = \gamma_{010}$

$\beta_{02j} = \gamma_{020}$

$\beta_{03j} = \gamma_{020} + \gamma_{021}\,(\text{DAP}) + u_{02j}$

$\beta_{04j} = \gamma_{040}$

$\beta_{05j} = \gamma_{050}$

$\beta_{06j} = \gamma_{060}$

$\beta_{07j} = \gamma_{070}$

$\beta_{10j} = \gamma_{100j}$

Fixed effect results of the fully conditional model are summarized in Table 40. According to findings, the predicted reading achievement at the end of kindergarten for the typical student (who was white, female, and average with respect to the AGE and SES) was found to be 27.62, in a classroom that was average on DAP score. The effect of DAP at this level was negative but statistically not significant on children's reading IRT scores. This nonsignificant effect could be explained as for each 10% increment in DAP score, the expected final status was reduced by 0.6 points (i.e., 10 X γ_{001}). After controlling for all student characteristics, all student level variables were still statistically significant. At the final data collection point, the effect of child SES, γ_{030}, was 4.32 points, indicating that an increase of one standard deviation in SES boosted the reading achievement 4.32 points, controlling for other variables in the model. Although it was not statistically significant, reading performance at the final stage for high SES students, however, declined slightly as a function of DAP (γ_{031}= -.03), indicating that higher SES children scored .03 units less in higher DAP classrooms.

Table 40
Fully Conditional Model for the Kindergarten Reading IRT Data

Fixed Effect	Coefficient	se	*t* ratio
Model for final status, π_{0ij}			
Model for Mean Status of a typical child β_{00j}			
Intercept γ_{000}	27.62*	.10	280.26
DAP γ_{001}	-.06	.04	-1.36
Model for MALE on final status β_{01j}			
intercept γ_{010}	-1.69*	.12	-13.71
Model for AGE on final status β_{02j}			
intercept γ_{020}	.37*	.02	21.50
Model for SES on final status β_{03j}			
intercept γ_{030}	4.32*	.12	35.31
DAP γ_{031}	-.03	.05	-0.66
Model for BLACK on final status β_{04j}			
intercept γ_{040}	-1.45*	.21	-6.98
Model for HISPANIC on final status β_{05j}			
intercept γ_{050}	-1.29*	.20	-6.35
Model for ASIAN on final status β_{06j}			
intercept γ_{060}	2.20*	.38	5.73
Model for OTHER on final status β_{07j}			
intercept γ_{070}	-1.82*	.37	-4.96
Model for growth rate, π_{1ij}			
Intercept γ_{10j}			
Intercept γ_{100}	.03*	.00	118.77

Estimated variances and related chi squared statistics from three level of the fully conditional model ware presented in Table 41. These results suggested that residual parameter variances still remained unexplained in π_{0ij}, $\beta_{00j,\text{ and}}\beta_{02j}$.

.

Table 41
Variance Decomposition from a three level analysis of the effects of selected student level predictors and DAP on reading performance

Random Effect	Variance	*df*	χ^2	*p*-value
Level 1				
Temporal variation, e_{tij}	2.86			
Level 2 (students within classrooms)				
Individual Final status r_{0ij}	47.30	12940	59871.29	.001
Level 3 (between classrooms)				
Classroom mean status, u_{00j}	3.45	944	1661.47	.001
Classroom mean status/SES u_{03j}	3.30	944	1371.43	.001

Kindergarten ARS Data

Fully Unconditional Model (Null Model)

Level -1 Model: At level-1 with an individual growth rate of reading achievement at time *t* of student *i* in classroom *j*:

$$\text{Reading (ARS)}_{tij} = \pi_{0ij} + \pi_{1ij}(\text{TIME})_{tij} + e_{tij}$$

Unconditional Level 2 (Student level)

$\pi_{0ij} = \beta_{00j} + r_{0ij}$

$\pi_{1ij} = \beta_{10j} + r_{1ij}$

Unconditional Level 3 (Classroom Level)

$\beta_{00j} = \gamma_{000} + u_{00j}$

$\beta_{10j} = \gamma_{100} + u_{10j}$

The findings from the fully unconditional model shown in Table 42, did not indicate a strong overall growth trajectory averaged across all children and classrooms. The estimated mean intercept, γ_{000}, was 2.98. The average reading learning rate within two semesters from Fall to Spring was estimated at .01 (γ_{100}), indicating that the average reading performance rate was statistically different from zero (p<.001).

Table 42

Three-Level Analysis of Kindergarten ARS Reading Data (Fully Unconditional Model)

Fixed Effect		Coefficient	se	*t* ratio
Average final status, γ_{000}		2.98	.011	267.98
Average growth rate, γ_{100}		.01	.001	69.34
Random Effect	Variance	*df*	χ^2	*p*-value
Level 1				
Temporal variation, e_{tij}	.093			
Level 2 (students within classrooms)				
Individual Final status r_{0ij}	.165	13267	18568.80	.001
Individual Growth rate, r_{1ij}	.001	13267	4767.28	>.500
Level 3 (between classrooms)				
Classroom mean status, u_{ooj}	.068	960	2708.14	.001
Classroom mean growth rate, u_{1oj}	.001	960	1133.51	.001
Level 1 Coefficient	Percentage of Variance Between Classrooms			
Final status, π_{0i}	.29			
Growth rate, π_{1i}	–			

Variance Covariance Components and Correlations among the Level-2 and Level-3 Random Effects

Level 2 $\begin{pmatrix} .165 & .854 \\ .000 & .00 \end{pmatrix} = \hat{\mathrm{T}}\pi = \begin{pmatrix} \hat{\tau}_{\pi 11} & \\ \hat{\tau}_{\pi 12} & \hat{\tau}_{\pi 22} \end{pmatrix}$

Level 3 $\begin{pmatrix} .068 & -.230 \\ -.000 & .001 \end{pmatrix} = \hat{\mathrm{T}}\beta = \begin{pmatrix} \hat{\tau}_{\beta 11} & \\ \hat{\tau}_{\beta 12} & \hat{\tau}_{\beta 22} \end{pmatrix}$

Note: the lower triangles contain the covariances; the upper triangles contain the correlations
**p<.05*

The decomposition of the variance in π_{0ij} and π_{1ij} into their within and between classroom components were examined and the estimates for these variance components are provided in Table 42. The chi-square statistics associated with these variance components indicated significant variation among students within classrooms for final status (π_{0ij}) and significant variation between classrooms for mean final status (β_{00j}). However, there was no evidence that reading growth rate (π_{1ij}) varied among students within classrooms for (p>.500) and also mean reading learning growth rates (β_{10j}) varied between classrooms (p>.500). That is, hypotheses $\hat{\tau}_{\pi 22} = \hat{V}ar(r_{1ij}) = 0$ and $\hat{\tau}_{\beta 22} = \hat{V}ar(u_{1ij}) = 0$ were sustained.

Based on these variance component estimates, the percentage of variation was computed as falling between classrooms for both final status and growth rates.

$$\text{\% Variance between classrooms on } \pi_{pjk} = \frac{\tau_{\beta pp}}{\tau_{\beta pp} + \tau_{\pi pp}}$$

Substituting corresponding estimates for the variance components into above equation yields the results [as following] presented in the third part of the Table 42;

% Variance between classrooms on π_{0ij} = (.068) / (.068+.165)= .29

% Variance between classrooms on π_{1ij} = undefined

Therefore, 29 % of the variance in final status lies between classrooms while no variance was observed between classrooms for the growth rates. The correlation between final status and growth rate into its within (level 2) and between-classroom (level 3) components were reported in Table 42 (See the last panel of the table). Within-and between classrooms, the estimated correlations were .85 and -.23, respectively.

In general, the variance component decomposition indicated an important feature of the kindergarten ARS data as the high percentage of the variation in reading learning rates laid within classrooms. The reliability estimates also tell us to determine whether certain random effects in within-and-between classrooms models might be constrained to zero. Estimated reliabilities for π_{0ij} (.33) and β_{00j} (.57) were evidently higher than for the π_{1ij} (.001) and β_{10j} (.20). Therefore, random effects of growth rates in reading among students and classrooms were fixed to be zero in the next models.

Student Model (Conditional Level 2 and Unconditional Level 3 Model)

In the first conditional model, student level variables were added to the model that allowed estimation of the effects of these student characteristics on individual reading achievement. The level 1 model remained same.

Level 1

Reading (ARS) $_{tij} = \pi_{0ij} + \pi_{1ij}(\text{TIME})_{tij} + e_{tij}$

The level 2 model represented the variability in final status among students within classrooms. Therefore, the random effect of growth rates among students was fixed in the second level based on the results of the fully unconditional model, indicating that there was no variation among students within classrooms in terms of reading growth rate. The level–2 model was formulated as follows. There was only one random effect within classroom (r_{0ij})

Conditional Level 2 (Student Level)

$$\pi_{0ij} = \beta_{00j} + \beta_{01j}(\text{MALE}) + \beta_{02j}(\text{AGE}) + \beta_{03j}(\text{SES}) + \beta_{04j}(\text{BLACK}) + \beta_{05j}(\text{HISPANIC}) + \beta_{06j}(\text{ASIAN}) + \beta_{07j}(\text{OTHER}) + r_{0ij}$$

However, the level-3 was still unconditional that means that it did not include any classroom level predictors at this time. Random effects of mean reading performance and the mean SES between classrooms were freely estimated but, remaining random effects associated with average student characteristics were fixed to zero based on results of the preliminary analysis. Then the Level-3 was modeled;

Unconditional Level 3 (Classroom Level)

$\beta_{00j} = \gamma_{000} + u_{00j}$
$\beta_{01j} = \gamma_{010}$
$\beta_{02j} = \gamma_{020}$
$\beta_{03j} = \gamma_{020} + u_{02j}$
$\beta_{04j} = \gamma_{040}$
$\beta_{05j} = \gamma_{050}$

$\beta_{06j} = \gamma_{060}$
$\beta_{07j} = \gamma_{070}$
$\beta_{10j} = \gamma_{100j} + u_{10j}$

The estimated fixed effects for this model are presented in Table 43. High magnitude *t*-statistics indicated that there were significant relationships between all student characteristics and the final reading status. Estimated mean reading achievement at the end of program for an individual who was average with respect to student predictors within a classroom was 3.01. The γ_{000} coefficient represented the predicted final status for a typical child (Female=0, White=0, average respect to the AGE and SES) in this study. For such a student, the predicted reading achievement at the end of the kindergarten was 3.01. Overall mean growth rate of reading achievement during the program was .01. In other words, from fall to spring semesters at the kindergarten level, children reading achievement would increase .01 points.

Table 43
Effects of Student-Level Characteristics on Student Kindergarten Reading Performance (ARS) (Conditional Level-2 Model)

Fixed Effect	Coefficient	se	*t* ratio
Model for final status, π_{0ij}			
Model for mean status of a typical child, β_{00j}			
Intercept γ_{000}	3.01*	.009	317.60
Model for MALE on final status β_{01j}			
Intercept γ_{010}	-.18*	.010	-17.53
Model for AGE on final status β_{02j}			
Intercept γ_{020}	.03*	.001	24.87
Model for SES on final status β_{03j}			
Intercept γ_{030}	.30*	.009	31.48
Model for BLACK on final status β_{04j}			
Intercept γ_{040}	-.11*	.020	-6.39
Model for HISPANIC on final status β_{051j}			
Intercept γ_{050}	-.21*	.019	-13.19
Model for ASIAN on final status β_{06j}			
Intercept γ_{060}	-.09*	..026	-3.81
Model for OTHER on final status β_{07j}			
Intercept γ_{070}	-.15*	.029	-5.58
Model for growth rate, π_{1ij}			
Intercept β_{10j}			
Intercept γ_{100}	.01*	.001	68.61

*p<.05

Within-classroom effect of MALE on the estimated mean reading achievement at the end of program for an individual was found to be -.18 (γ_{010}), indicating that male students finished the program .18 points behind their female counterparts (classroom mates). One standard deviation increase in age was associated with .03 (γ_{020}) higher reading scores, indicating an advantage for older students over young students in kindergarten. The effect of SES on the reading performance at the end of program for an individual was found to be .30 (γ_{030}), indicating that a student with one standard deviation higher SES level ended the program .30 points higher than others. The effect of BLACK at the end of program for an individual was found to be -.11 (γ_{040}), demonstrating that an African-American student finished the program .11 points behind his or her Caucasian counterparts within classroom *j*. The effect of HISPANIC on the ARS reading scores at the end of program for an individual was found to be -.21 (γ_{050}), indicating that a Hispanic student finished the program .21 points behind Caucasian students within classroom *j*. Unlike the effects observed in reading IRT score models, the effect of ASIAN on the ARS reading achievement at the end of program for an individual was found to be -.09 (γ_{060}), indicating that an Asian student finished the program .09 points behind Caucasian students within classroom *j*. The effect of OTHER on the reading performance at the end of program for an individual was found to be -.15 (γ_{060}), indicating that a student who was in the other race group rather than white, black, Hispanic, or Asian, finished the program .15 points behind his or her counterparts within classroom *j*.

Estimated variances and related chi squared statistics from the three-level decomposition are presented in Table 44. These results suggest that residual parameter variances still remained to be unexplained in π_{0ij}, β_{00j}, and β_{02j}.

Table 44
Variance Decomposition from a three level analysis of the effects of selected student level variables on student reading learning

Random Effect	Variance	*df*	χ^2	*p*-value
Level 1				
Temporal variation e_{tij}	.025			
Level 2 (students within classrooms)	.264	12303	39090.11	.001
Individual Final status r_{0ij}				
Level 3 (between classrooms)				
Classroom mean status u_{00j}	.046	942	2163.13	.001
Classroom mean status/SES u_{03j}	.021	942	1310.62	.001
Classroom mean growth rate u_{10j}	.001	942	4066.90	.001

Classroom model (Fully conditional model)

For this model, a classroom level continuous variable "DAP" was added to model to examine the effect of DAP on students' final reading performance and the effect of the relationship between SES and DAP on students' final reading performance.

Conditional Level 2 (Student Level)

$\pi_{0ij} = \beta_{00j} + \beta_{01j}\text{(MALE)} + \beta_{02j}\text{(AGE)} + \beta_{03j}\text{(SES)} + \beta_{04j}\text{(BLACK)} +$

$\beta_{05j}\text{(HISPANIC)} + \beta_{06j}\text{(ASIAN)} + \beta_{07j}\text{(OTHER)}\ r_{0ij}$

$\pi_{1ij} = \beta_{10j}$

Conditional Level 3 (Classroom Level)

$\beta_{00j} = \gamma_{000} + \gamma_{001}\text{(DAP)} + u_{00j}$

$\beta_{01j} = \gamma_{010}$

$\beta_{02j} = \gamma_{020}$

$\beta_{03j} = \gamma_{020} + \gamma_{021}\text{(DAP)} + u_{02j}$

$\beta_{04j} = \gamma_{040}$

$\beta_{05j} = \gamma_{050}$

$\beta_{06j} = \gamma_{060}$

$\beta_{07j} = \gamma_{070}$

$\beta_{10j} = \gamma_{100j} + \gamma_{101}\text{(DAP)} + u_{10j}$

Fixed effect results of the fully conditional model are demonstrated in Table 45. The findings revealed that reading achievement score at the end of the kindergarten, γ_{000}, was 3.00 for the typical student who was white, female, and average with respect to the AGE and SES. The effect of DAP was statistically significant, unlike the insignificant effect observed in reading IRT scores. For each 10% increment in DAP scores, the expected final status was increased by 0.1 points (i.e., 10 X γ_{001}). The effects of student level variables displayed a similar pattern to the previous model (conditional level 2 and unconditional level 3), as they were all significant and comparable in magnitude. At the final data collection point, the effect of child SES, γ_{030}, was .30. This means that an increase of one standard deviation in SES boosted the reading achievement .30 points, referring to 10% of the total reading achievement (γ_{000}=3.00). Furthermore, the effect of DAP on the effect of SES was statistically significant and positive in direction (γ_{031}= .01, t= 2.39, p= .017). Reading performance at the final stage for students with higher SES levels increased significantly as a function of DAP. However, the effect of DAP on the growth rate in kindergarten year was not significant (γ_{101}= .01, t= 0.87), yet positive.

Table 45
Fully Conditional Model for the Kindergarten Reading ARS Data

Fixed Effect	Coefficient	se	*t* ratio
Model for final status, π_{0ij}			
Model for Mean Status of a typical child β_{00j}			
Intercept γ_{000}	3.00	0.009	316.71
DAP γ_{001}	0.01	0.003	2.39
Model for MALE on final status β_{01j}			
Intercept γ_{010}	-0.18	0.010	-17.56
Model for AGE on final status β_{02j}			
Intercept γ_{020}	0.03	0.001	24.87
Model for SES on final status β_{03j}			
Intercept γ_{030}	0.30	0.010	30.77
DAP γ_{031}	0.01	0.003	2.34
Model for BLACK on final status β_{04j}			
Intercept γ_{040}	-0.10	0.017	-6.29
Model for HISPANIC on final status β_{05j}			
Intercept γ_{050}	-0.21	0.016	-13.13
Model for ASIAN on final status β_{06j}			
Intercept γ_{060}	-0.09	0.025	-3.76
Model for OTHER on final status β_{07j}			
Intercept γ_{070}	-0.15	0.027	-5.56
Model for growth rate, π_{1ij}			
Intercept β_{10j}			
Intercept γ_{100}	0.01	0.001	68.54
DAP γ_{101}	0.01	0.001	0.87

Estimated variances and related chi squared statistics from three level of the fully conditional model are presented in Table 46. These results also suggest that residual parameter variances still remained unexplained in all levels, π_{0ij}, β_{00j}, and β_{02j}, respectively.

Table 46
Variance Decomposition from a three level analysis of the effects of selected student level predictors and DAP on reading performance

Random Effect	Variance Component	*df*	χ^2	*p*-value
Level 1				
Temporal variation, e_{tij}	.025			
Level 2 (students within classrooms)				
Individual Final status r_{0ij}	.264	12363	39199.69	.001
Level 3 (between classrooms)				
Classroom mean status, u_{00j}	.046	941	2151.88	.001
Classroom mean status/SES u_{03j}	.021	941	1303.31	.001
Classroom mean growth rate, u_{10j}	.001	941	4061.67	.001

The following table summarizes the major findings on both measures of reading. The most powerful pattern in the examination of the findings was observed on the effects of DAP on two different measure of DAP. The effects of DAP were nonsignificant on the standardized measure of reading achievement, whereas the same effects were significant on all nonstandardized indirect measures. The findings did not reveal a convincing pattern of differential effects for DAP across various groups of children.

Table 47.
Summary of Estimated Findings on Main Variables

Estimated Effects	Standardized Direct Measure			Indirect Measure of ARS		
	K	1st	3rd	K	1st	3rd
Main effect of DAP	NS	NS	NS	Positive	Positive	Positive
Interaction with DAP						
SES	NS	NS	NS	Positive	NS	NS
Age	NS	NS	NS	NS	NS	NS
Gender	NS	NS	NS	NS	NS	NS
Race	NS	NS	Positive (African-American)	NS	NS	Negative (Hispanic)
Sector	Positive (Public)	NS	NS	NS	Positive (Public)	NS

Note: NS = Non-Significant effect.

CHAPTER V

DISCUSSION

This chapter presents a summary of the study, major findings, and a discussion of the implications and directions for future research. The chapter starts with a brief summary of the study, where the purposes, the theoretical background that led to the research questions, and the procedure are discussed. Major findings of the study are provided in the next section. Then, the major findings are discussed with reference to both the theoretical conceptualization of developmentally appropriate practices and results of previous research. Finally, implications that can be drawn from the study and recommendations for future research are presented.

The purpose of the study was to examine the effects of developmentally appropriate practices on children's academic achievement, namely reading scores, from kindergarten through the first three years of elementary education. Limitations of previous research methodology shaped the decision to employ an appropriate multilevel analysis approach that considers the nested data structure. Furthermore, the differential effects of DAP as a function of children's age, gender, family SES, ethnicity, and school sector were examined.

Since the NAEYC's position statement on developmentally appropriate practices was first published in 1987 (Bredekamp, 1987), the concept of developmental appropriateness has broadly affected both practice and policy in the field of early childhood education. Arguably, DAP statement has become one of the most influential documents guiding the field of early childhood education (Hart, Burts, & Charlesworth, 1997). Developmentally appropriate practice (Bredekamp, 1987; Bredekamp & Copple, 1997) was introduced by the National Association for the Education of Young Children to illustrate the best possible practices, environments, and pedagogies for young children. The philosophical underpinnings of DAP were influenced by philosophical and educational traditions ranging from Jean Jacques Rousseau's romanticism to Piagetian constructivism and maturationist psychology. Developmentally appropriate practice is predominantly based on the theories of Piaget (1952), Erikson (1963), and Vygotsky (1978). Based on Piaget's idea of predictable sequences of growth and change, it was claimed that knowledge of typical development of children within the age span grants a general framework to organize the learning environment, curriculum goals, and appropriate practices (Bredekamp & Copple, 1997). Furthermore, NAEYC's position statement initiates that children are active learners and they are "…actively engaged in constructing their own understanding from their experiences, and these understandings are mediated by and clearly linked to the sociocultural context. Young children actively learn from observing and participating with other children and adults…" (Bredekamp & Copple, 1997, p. 13).

Vygotsky's sociocultural theory was also at the center of the philosophy of DAP. According to Vygotskian theory, learning occurs when children interact with both people and materials in their environments. The social cognition learning model views the culture as the primary determinant of individual development. Therefore, a child's learning and development is affected by the surrounding culture, including the culture of family environment in which he or she is enmeshed. Vygotsky's theory of the zone of proximal development (Vygotsky, 1978) informed the relationship between the teacher and the child in DAP classrooms. Teachers provide a variety of activities and materials as well as promoting collaboration with other children. Teachers also adjust the level of difficulty of an activity, and in doing so, provide assistance to support the child's level of performance.

Although the concept of DAP has been very popular and influential in the field of early childhood education, the number of empirical studies evaluating the effects of DAP have been quite limited (Jones & Gullo, 1999; Van Horn & Ramey, 2003). A review of the literature revealed that the results of the existing studies are weak and mixed. Of those studies investigating the effects of DAP on the cognitive domain, five studies (Burts et al., 1993; Dunn, Beach, & Kontos, 1994; Marcon, 1992, 1993, 1999) found positive effects, five studies reported mixed or no effects (Hirsh-Pasek et al., 1990; Huffman & Speer, 2000; Jones & Gullo, 1999; Stipek et al., 1998; Van Horn & Ramey, 2003), and one study reported negative effects (Stipek et al., 1995). There was also another issue that these studies were fairly limited in sample sizes; that is 37 to 307 children in 2 to 165 classrooms were included in those studies. Furthermore, only four studies took place in kindergarten, one in both kindergarten and first grade, and one in first grade. Another methodological limitation was revealed after the literature review that most of the studies investigating the effects of DAP did not consider the nature of the nested data, where children are nested within classrooms which in return are nested within schools. Therefore, an appropriate multi level data analysis approach was utilized in this study.

The data for the multilevel analyses was drawn from the Early Childhood Longitudinal Study-Kindergarten Class of 1998-99 (ECLS-K), a “multi-source, multi-method study that focuses on children’s early school experiences beginning with kindergarten” (US Department of Education, 2000, p, 1-1). The ECLS-K provided data from a nationally representative sample of children from kindergarten through third grade. A total of 22,782 children throughout the US were sampled in the study and assessed directly or indirectly in 1,277 schools which offered kindergarten programs during the 1998-99 school year. It was expected that the ECLS-K data set would provide a more thorough description of DAP and, in turn, a more accurate estimate of the effects of DAP on academic achievement. This is because, when used with appropriate sample weights, results from the ECLS-K data are generalizable to the United States’ population of kindergarten children, teachers, and schools offering kindergarten programs in the 1998-1999 school year. Considering the fact that only one study used what can be considered a large sample (Van Horn & Ramey, 2003), the current study was thought to add significant information to the literature by utilizing a nationally representative ECLS-K data.

This study employed four waves of data from the ECLS-K; the fall of 1998, spring of 1999, spring of 2000, and spring of 2002. Two attempts were made to filter the available sample of 22,782 kindergarten children in 1,277 schools. To start with, only first time kindergartners who attended kindergarten in 1998-1999 were included in this study. In addition, only children who advanced the first grade and the third grade in the normal pace were selected. Therefore, children who repeated either one of the kindergarten, the first grade, or the third grade were excluded from the study. The final sample is representative of children who attended kindergarten for the first time in the school year of 1998-1999. Next, children who had both direct and indirect measures of reading, as well as measures of DAP, at four time points were included in this study. The final sample of this study consisted of 14,398 children in the kindergarten year, 9,677 children in the first grade, and 8,149 children in the third grade.

The outcome variables of interest were one direct and one indirect measure of children's reading achievement. Because standardized tests were considered to be inconsistent with the philosophical basis of developmentally appropriate practices, it was thought that measuring the effects of DAP on a process-based indirect measure would provide additional information. Therefore, the indirect measure of reading ARS scores (Academic Rating Scale), in addition to the direct measure of reading IRT scores, were utilized to alternatively assess the effects of DAP. Other variables included in the study were student characteristics of age, SES, race/ethnicity, gender, and the school level variable of sector (public vs. private).

The next section presents a discussion of the findings. The section starts with a presentation of the findings regarding the child level demographics of gender, SES, age, and race/ethnicity. This part is followed by a discussion of the main findings in relation to the effects of DAP.

Students' Reading Achievement from Kindergarten through Third Grade

In six multilevel cross-sectional analyses, the effects of student level characteristics of gender, SES, age, and race/ethnicity on two measures of reading achievement were investigated. The results indicated that inclusion of these student level variables, along with DAP and the school level indicator of sector reduced the school level variance in the standardized measure of reading achievement by 17%, 13%, and 7% in the kindergarten, the first grade, and the third grade, respectively. The same variables reduced the school level variance in the nonstandardized indirect measure of reading achievement by 17% in both the kindergarten and the first grade, and by 25% in the third grade. These variables seemed to be better predictors of reading achievement when it was assessed by a nonstandardized measure.

The three-level cross-sectional analyses illustrated the national status of children's reading performance. According to theses analyses, of all of the variables included in the models, socioeconomic status (SES) had the most substantial effect on students' reading achievement, in both the direct and indirect measures. This statistically significant effect was consistent across all three grades. Higher SES children achieved better reading scores on both measures, after controlling for other variables. These results were consistent with previous reports of NCES (2001). However, the magnitude of the effect of SES was slightly larger on the indirect measure of reading ARS scores. A one point increase in SES was associated with almost 35% of a standard deviation (SD) increase in the ARS scores, whereas the effect was about 32% of a SD on the standardized reading IRT scores, on average. Specifically, the effect of SES on reading IRT scores increased through the grades, and it showed the largest effect in the third grade.

The next influential student level predictor of reading achievement was race/ethnicity. In general, Asian children scored higher than other children on the reading IRT measure in both kindergarten and the first grade, after controlling for the other variables in the model. However, the effect for Asian children was statistically non-significant and negative in magnitude in the third grade. Furthermore, the advantage for Asian children was not repeated on the reading ARS scores. This variation in the effects might be explained by the fact that the ARS scores were teacher-reported indirect measures of reading. The Asian sample in this study is likely to include children who are second language speakers of English and this might have affected the rating of teachers. Caucasian children were the next highest achievers on both measures across all grades.

The most disadvantaged children on both measures were African-American, followed by the Hispanic children, and children grouped as "Other" (Pacific Islanders, Native Americans, and others). Unlike other races, the gap between the African-American children and others widened throughout grades, showing the largest effect in the third grade. In the third grade, African-American students scored 53% of a SD lower than all other students, after controlling for the variables in the model. This finding mimics the effect of SES, which also had reached its' peak in the third grade. Considering the historically well-represented high correlation between SES and race, the third grade appears to be a key point in children's school success. The gap between African-American children and low SES children widens in the third grade. Further analyses of this relationship in later grades seem necessary for consideration of policy issues in the field. The effect of the gender on students' reading achievement was consistent across grade on both measures. After controlling for SES and other variables, girls consistently outperformed their boy counterparts across all three grades. Girls had a 2-point advantage on reading IRT scores in kindergarten and the first grade and a 3-point advantage in the third grade, referring to approximately 18% of a SD. These findings were consistent with the results reported in previous research on gender gap (Sheppard, 1986; NCES, 2001), as girls generally develop faster than boys. The least influential of variables included was the age of children at the time of the measurement. Although statistically significant, the effect of the age of a child was trivial, showing a difference of less than 5% of a SD. This statistically significant, but practically trivial, effect indicated that older children had better reading achievement on both measures, after controlling for the other variables in the model.

The Effects of Developmentally Appropriate Practices on Reading Achievement

The effects of DAP on children's reading achievement was addressed through multilevel cross-sectional analyses over three grades and multilevel growth analyses in the kindergarten year. Furthermore, the hypothesis that DAP had differential effects across some student and school characteristics was tested by including age, gender, SES, race/ethnicity, and school sector in the models.

The findings indicated that DAP did not have a significant effect on students' reading achievement as measured by the standardized direct measure, IRT scores. This non-significant effect was consistent across the grades, as well as the growth analysis in kindergarten. Furthermore, the effects were positive in direction in kindergarten and the third grade, but negative in the first grade. The magnitudes of the effects were also very small, almost nonexistent. Considering the fact that this data was drawn from a nationally representative sample, it raises questions against a widely applied concept of DAP in school settings. The results of no significant effects mimic the findings reported by the most recent study of Van Horn and Ramey (2003). The study by Van Horn and Ramey (2003) is of great importance because of the fact that their study addressed the limitations of previous research and used an appropriate multilevel analysis approach. It provided a rare example of adequate sample size and statistical analysis in DAP research; the sample included 4764 children in 1537 classrooms, and a multilevel statistical analysis was employed. As the most recent study of DAP, they did not find any significant effect of DAP in any grade level and growth over time on any of the standardized outcome measures. Similarities in the methodology used and the large sample sizes utilized in both studies make the argument that DAP has no effects on students' standardized reading achievement from kindergarten through the third grade more viable. Furthermore, this result is consistent with another group of previous studies where DAP was not found to be related to student achievement on cognitive outcomes (i.e., Hirsh-Pasek et al., 1990; Huffman & Speer, 2000; Jones & Gullo, 1999; Stipek et al., 1998). A plausible explanation for no effects can be found in the fact that the current study employed more control variables that previous research that found positive effects. Research on the effects of DAP generally examined it in isolation (Van Horn, Karlin, Ramey, Aldridge, & Snyder, 2005). Few studies included one or two child level control variables like gender and ethnicity (e.g., Harts et al., 1998; Marcon, 1993). The current study is the first that has attempted to examine the effects of DAP in relation to the four child characteristics: gender, SES, age, and race/ethnicity. This can explain the lack of evidence for significant effects on the standardized measure of academic reading achievement. Possibly, the effects of DAP may be not be apparent with the inclusion of variables such as age or SES, which are highly related to the developmental level of a child.

There was, however, a different picture when examining the effects of DAP on the nonstandardized indirect measure of reading. The Academic Rating Scale (ARS) was developed to assess the same set of skills of reading with a different approach that was more process-oriented, rather than being product-oriented. It indirectly measured the reading skills of children by teacher ratings. Results were completely in contrast with the findings on standardized IRT scores. DAP was found to be significantly related to children's reading achievement as measured by the ARS across all grades. This interesting finding raises important issues about how to assess the effects of DAP. Although there were no effects of DAP on the standardized reading test score, it showed significant effects on a nonstandardized, informal measure, of reading score. This finding requires a special attention because DAP is one of the most influential concepts, if not the most influential, in shaping instructional practices in the field.

The direct measure of reading is a standardized test that was developed by borrowing or adapting items from published tests including the Peabody Individual Achievement Test-Revised (PIAT-R), Peabody Picture Vocabulary Test-Revised (PPVT-R), the Primary Test of Cognitive Skills (PTCS), the Test of Early Reading Ability (TERA-2), the Test of Early Mathematics Ability (TEMA-2), and the Woodcock-Johnson Tests of Achievement-Revised (WJ-R) (Rock & Pollack, 2002; Rathbun, West, & Germino-Hausken, 2004). On the other hand, the ARS was an indirect measure of teachers' evaluations of student academic performance. Although both direct and indirect assessments were designed to collect information about children's academic skills and knowledge on the exact same subject areas, there were some variations between these two measures.

The ARS mostly focused on both the processes and products of children's learning, whereas, due to time and space limitations, the direct cognitive assessments measured only the products of children's learning. Constraints of standardized testing in direct cognitive measures did not limit the development of the ARS items. The ARS was designed to include skills, knowledge, and behaviors that reflect a broader sampling of the most recent national curriculum standards and guidelines from early childhood professionals and researchers.

The first issue to address here is whether teachers can be valid assessors of their students' academic performance as the ARS was a teacher-reported measure. The ARS was a measure of reading that mostly focused on both the process and products of children's learning, whereas, due to time and space limitations, the direct cognitive assessments measured only the products of children's learning. Consequently, the data for the outcome measures of ARS reading scores were derived from the spring semester of each grade level, meaning that teachers had sufficient amount of time to reliably observe the abilities of their students during a complete school year. Researchers in the field have argued that because teachers have a chance to observe and interact with students on a daily basis, they are capable of validly assessing their students' performance, and they are in the best position to make decisions about their students' intellectual, socio-emotional, and behavioral accomplishments (e.g., Hopkins, George, & Williams, 1985; Meisels, Bickel, Nicholson, Xue, & Atkins-Burnett, 2001; Perry &

Meisels, 1996; Kenny & Chekaluk, 1993). Some other researchers, on the other hand, voiced their concerns about the validity and reliability of these assessments (Hoge & Coladarci, 1989). Teachers' judgments on student performance are thought to be possibly affected by their expectations and biases (Hoge, 1983; Hoge & Butcher, 1984; Sharpley & Edgar, 1986; Silverstein, Brownlee, Legutki, & MacMillan, 1983).

Regarding the concerns of validity of teacher judgments, studies by Hoge and Caladarci (1989) and Hoge and Butcher (1984) provided evidence of criterion validity. Hoge and Coladarci (1989) reviewed a total of 17 studies of teacher judgments and reported a median correlation of .66 between teacher ratings and children's achievement level on standardized tests. In two different studies, Coladarci (1986) and Leinhardt (1983) investigated the accuracy of teachers' judgments about whether their students would correctly answer academic ability questions on an item by item analysis. They reported that teachers' judgments were accurate 70% and 64% of the times, respectively. Moreover, the accuracy of teacher judgments in predicting student's future performance is also well documented (Salvesen & Undheim, 1994; Stevenson, Parker, Wilkinson, Hegion, & Fish, 1976). Stevenson et al. (1976) reported that teachers' judgments were able to predict students' academic performance in reading and mathematics from kindergarten through the third grade. However, the accuracy of teacher ratings was different across subject areas. Hoge and Coladarci (1989) indicated that the accuracy of teacher judgments ranged between $r = .28$ and $r = .92$. Teachers' judgments were more accurate in reading, language arts, and mathematics than social studies (Coladarci, 1986; Hopkins et al., 1985). Furthermore, teachers were better able to judge students' observable performance than less observable skills (Coladarci, 1986). Regarding the evidence for the content validity, Leinhardt and Seewald (1981) illustrated that teachers' judgments better reflected the content covered in the curriculum, because they accounted for the classroom instruction. In light of the discussion above, utilizing the ARS measure in the ECLS-K data to assess students' academic reading achievement is well justified.

Another concern to be addressed is the reliability of teacher judgments in general, and the reliability of the ARS reading scores in this study specifically. Evidence for the test-retest reliability (e.g., Stevenson et al., 1976), inter-rater reliability (e.g., Hoge, 1993), and internal consistency (Meisels et al., 1995) is available in the literature. The study by Meisels et al. (2001) reported an inter-rater reliability of r = .88 and internal consistency Alphas ranging from .87 to .94. The reliability levels reported by the NCES for the reading ARS scale are also very high. Reliability estimates for the ARS Language and Literacy scores used in this study were .87, .91, .94, and .95, for the fall kindergarten, spring kindergarten, spring first-grade, and spring third-grade, respectively.

Returning to the question of whether the ARS is a valid and reliable measure of children's reading achievement, it is believed that the findings of previous research, and the psychometric indicators reported by the NCES support the use of the ARS to alternatively assess the academic reading achievement of children. The report by Perry and Meisels (1996) was prepared before designing the ARS measure for the ECLS-K. After examining the national and state standards, as well as the literature on the predictive validity of early academic skills, recommendations made by them guided the efforts of developing meaningful and trustworthy measures of teacher judgments, namely the ARS. Depending on Perry and Meisels' (1996) recommendations, and the available literature, items in the ARS measures were highly specific and teachers were provided with clear examples to distinguish between the levels of difficulty of an item on the ARS in order to reach a satisfactory level of specificity and sensitivity. Furthermore, The ARS included some criterion-referenced items with high level of specificity to eliminate teacher bias based on the findings of previous research on teacher judgments (Meisels et al.; 2001). The ARS also incorporated items "that overlap with the content assessed through the direct cognitive battery" and "some items that expand the skills tested by the direct cognitive battery—particularly those that assess process skills that would be difficult to assess given the time constraints" (Rock & Pollack, 2002, p. 2-18). Finally, specific to the reading ARS measure, some items aimed to assess the listening, speaking, reading, and writing skills. To sum up, review of the previous research and the psychometric reports by the NCES support the use of the ARS as a teacher-reported measure of students' academic achievement to collect data from kindergarten through the third grade.

Having concluded that the ARS was a reliable and valid measure of same reading skills as the standardized IRT-based direct measure, emerges a subsequent issue of how the effects of DAP should be assessed. In the present study, DAP showed significant and positive effects on the non-standardized indirect measure of the ARS, but showed no effects on the standardized IRT-based direct measure. This differential effect calls for caution in investigating the effects of DAP with various measures of achievement. The results of this study suggest that it would be more appropriate to measure the effects of DAP on academic achievement with alternative assessment methods, rather than using a standardized measure.

Using standardized tests to measure the effects of developmentally appropriate practices has long been criticized because the nature of DAP is thought to be contrary to the philosophy of standardized tests (e.g., Bredekamp, 1997; Hart, Burts, & Charlesworth, 1997; Kostelnik, Soderman, & Whiren, 2004; Miller, 1996). Some insignificant or no effects found in the literature were also explained by this contradiction, based on the fact that whereas DAP is more process oriented, standardized tests generally measure the final product. Consequently, measuring the effects of DAP with standardized tests considered to be inappropriate by some researchers in the field.

Standardized testing of young children has been a controversial issue since the trend of accountability in the early 1980's. In response to the 1983 report, *A Nation at Risk*, standardized tests had been increasingly utilized in education. In particular, using standardized test in the early years was highly criticized by scholars and various professional organizations (e.g., International Reading Association, 1986). Inappropriate assessment was viewed to be the inevitable outcome of the inappropriate curriculum (Sheppard, Taylor, & Kagan, 1996). Accordingly, the statement of DAP was followed by the position statement on standardized testing by the NAEYC (1988).

NAEYC emphasized the inappropriate use of standardized test in early childhood years and objected "to overuse, misuse, and abuse of formal, standardized testing, epitomized by standardized achievement tests that are unrelated to the ongoing activities of classrooms" (Bredekamp & Rosegrant, 1992, p. 44). The key in this statement is to find an appropriate assessment method that reflects the curriculum and that is more consistent with daily practices of a classroom. Accordingly, the guidelines of NAEYC on developmentally appropriate assessment call for involving "regular and periodic observation of the child in a wide variety of circumstances that are representative of the child's behavior in the program over time" (Bredekamp & Rosegrant, 1992, p. 23). In general, NAEYC encourages the utilization of observation-based informal methods assessments that are more appropriate to the nature and the developmental level of the child.

The reading ARS is a teacher reported measure that reflects the observations of the teacher on each child. Particularly, teachers had a great number of opportunities to observe the abilities of students when reporting the reading ARS scores in spring semesters. The ARS scores collected at the spring of kindergarten, the first grade, and the third grade illustrate the judgments of teachers that were developed during an entire semester and across many contexts. Teachers' judgments reflected in depth knowledge of students Therefore, it was concluded that the ARS provided a more developmentally appropriate assessment of children's reading achievement, than the standardized direct measure of reading. Still, although the literature did not reveal evidence for teacher bias, scores from the teacher rated ARS must be evaluated carefully (Perry & Meisels, 1996).

In light of the discussion above, future studies must consider the finding reported in this study that DAP did not show any significant effect on the standardized direct measure of reading, whereas the effects were significant and positive on the indirect and more developmentally appropriate measure. This differential effect of DAP must be taken into account in choosing the outcome measures and discussing the results regarding DAP.

The Differential Effects of Developmentally Appropriate Practices on Reading Achievement by Age, SES, Race, and School Sector

The results provided evidence for two differential effects of DAP on the direct standardized measure in kindergarten and the third grade, despite no main effects. Regarding the indirect measure of the ARS, DAP resulted in three differential effects, one in each grade.

In kindergarten, children in public schools benefited more from DAP on the standardized measure. On the non-standardized indirect measure, however, results indicated a significant interaction effect between DAP and SES, as higher SES children were rated higher in higher DAP classrooms. These results are controversial to the study conducted by Burts et al. (1992). Using a sample of 166 kindergarten children from a previous study (Burts et al., 1992), Burts et al. (1993) reported that low SES children had better overall GPAs in more DAP classrooms, whereas high SES children performed better in non-DAP classrooms. However, caution must be taken regarding the small sample size of that study, lack of appropriate multilevel research methodology, and the reliability and validity issues of GPA scores. Results from the current study are representative to the U.S population and they were generated by utilizing an appropriate hierarchical design that considered the nested nature of the data. However, it is also plausible to argue that a rating bias may have been reflected in the ratings of teachers in assessing high SES children.

In the first grade, there was only one differential effect of DAP that was observed on the non-standardized indirect measure of reading. Again, children in public schools, as opposed to the private schools, benefited more from DAP. Previous research fails to explain differing effects of DAP in public and private schools. Although there was a private school advantage over public schools on both reading measures at each grade, students in public schools had better scores as a function of DAP. This somewhat consistent relationship between two variables requires further investigation on various achievement scores.

The hypothesis that DAP has differential effects across various race groups was also scrutinized. The findings indicated that DAP had differential effects only on two race groups, both in the third grade. African-American students performed significantly better in higher DAP classrooms, assessed by the standardized direct measure. Considering the fact that African-American children scored over half a standard deviation less than other students without taking any classroom level variables into account, DAP tends to serve to decrease the level of disadvantage experienced by this minority group. A smaller differential effect was also observed on the non-standardized indirect measure of reading for the Hispanic race group as they scored lower in higher DAP classrooms. However, regarding the non-significant effects reported in previous studies (e.g., Burts et al., 1992; Van Horn & Ramey, 2003) and the large number of analyses in this study, extra care must be taken in interpreting the effects of DAP in relation to the other variables. In general, it can be concluded that the effects of DAP across various race groups did not systematically differ. Furthermore, no interaction effects of DAP across different age and genders were detected. The suggestion that girls perform worse than boys in higher DAP classrooms (e.g., Burts et al., 1992; Marcon, 1992, 1993) did not stand true, after controlling for other student and school level variables. No systematic differential effect of DAP across age and gender groups was revealed in the current study, Given the nationally representative sample and the appropriate design methodology, the argument about the bias in high DAP classrooms against certain groups of children was not supported.

Implications for Practice

The results of this study provide several possible implications to the practice of early childhood education. First, the measure of DAP in this study provided a valid measure of DAP. Given the evidence that the assessment of DAP has been a problematic issue in previous research (Van Horn & Ramey, 2003), the indicators of DAP included in this study should be considered in the curriculum design efforts and future research efforts when examining DAP. Assessing a concept as broad as DAP provides a great deal of challenge to researchers in the field. Although the NAEYC presented the guidelines in detail about what constitutes a developmentally appropriate program, it is still not clear how to assess DAP, and there is no one single instrument that is widely accepted as the best instrument in the field. In addition, it is complicated to assess some of the aspects of DAP like parental involvement. Therefore, in this study, three distinct aspects of DAP were employed: General classroom practices; availability of classroom centers; and availability of classroom materials.

Focusing on some of the main indicators of DAP, instead of trying to cover every possible one, might be more beneficial, as Maxwell and colleagues suggested (Maxwell, McWilliam, Hemmeter, Ault, & Schuster, 2001). This narrower approach may also serve in providing a better understanding of DAP for the general public. Because the positive effects of DAP were revealed, it is suggested that teachers incorporate the activities and materials that were used in this study to describe DAP. In this study, DAP was measured as a continuous variable. Of the studies that investigated the effects of DAP on child outcomes, all but one utilized an approach where DAP was created as a categorical variable with two or three categories. The most common practice was that DAP classrooms were compared to DIP (developmentally inappropriate practice) classrooms.

In some other studies, a third group that was placed neither DAP nor DIP but somewhere in between. However, both the NAEYC guidelines (Bredekamp, 1987; Bredekamp & Copple, 1997) as well as researchers from both sides of the debate concur that DAP exists on a *continuum*, as opposed to an *either/or* formation. The conceptualization of DAP as a continuous construct may contribute to future empirical research efforts.

The findings also revealed that girls and high SES children perform better in reading during the first four years of schooling. This situation might be used to support underachieving children in DAP classrooms, where the cultural differences are taken into account in daily classroom practices. Matching a more capable child with a low achiever in heterogeneous reading groups in DAP classrooms might foster both children's capabilities. DAP classrooms may provide a valuable environment that promotes social interaction between diverse groups of children that is culturally sensitive and individually appropriate to the level of each child (Bredekamp & Copple, 1997).

The findings provide empirical evidence for positive effects of DAP on children's early reading achievement when it is measured by a more developmentally appropriate and inclusive measure. However, DAP was not found to have significant effects on reading achievement in some previous research. Positive findings of this study raise questions about the instruments used to measure student achievement in those studies. The non-significant effects may have been due to the utilization of standardized measures in studies. When assessing the efficiency of DAP programs, policy makers should consider the fact that DAP may demonstrate significant effects on various measures of children's development. Measuring the success of DAP curriculum depending only on the results of standardized tests may be misleading in shaping practice in early childhood classrooms. Stake holders of the system must be aware that DAP may have differential effects on various measures, such as more process-oriented alternative methods of assessment. Such alternative ways of assessment should be included in future consideration of instructional practices in early childhood classrooms. The findings support the use of the assessment guidelines outlined in the position statement of NAEYC that assessment efforts in the early years must include developmentally appropriate methods (Bredekamp & Rosegrant, 1992). Developmentally appropriate assessments must be included in evaluating the accountability of early childhood initiatives, in addition to standardized testing, which has been widely utilized in the past two decades as the sole criterion of a program's success or failure.

Finally, the findings revealed that DAP did not hinder children's reading achievement from kindergarten through the third grade. Although there was not a positive effect of DAP on the standardized reading scores, there was no evidence of a negative effect either. Proponents of a more didactic and teacher-directed instruction in the early grades have argued that a child-initiated approach may not be as effective in developing children's academic skills, and that children from a child-initiated classroom do not perform at the level they could have, had they attended to a more teacher-directed program. Given that the findings from the current study did not indicate negative effects of DAP, even on standardized measures of academic performance, and there were significant positive effects of DAP on indirect measures of academic achievement, this study suggests that practitioners should adopt a more developmentally appropriate approach to early childhood education. Furthermore, evidence is available that DAP exhibited a consistent positive effect on many psychosocial measure including higher motivation (Stipek et al., 1995) and creativity (Hirsh-Pasek et al., 1990), as well as reduced stress behaviors (Hart et al., 1998). Therefore, the adoption of developmentally appropriate practices, as outlined in NAEYC's publication, is supported by the findings of the current study.

Recommendations for Future Research

In view of the limitations of previous research methodology, and the nature of the nested data structure, an appropriate multilevel analysis was employed to investigate the extent to which DAP is associated with children's development in reading achievement from kindergarten through third grade. Previous research, with the exception of the study conducted by Van Horn and Ramey (2003), was limited by either a small sample size or by a lack of appropriate multilevel research design that would fit the nested nature of the data. Failure to utilize a multilevel nested design with hierarchical data results in producing standard errors that are too small. In turn, this would increas the likelihood of finding a significant difference (Hox, 1995, 1998). This phenomenon may have contributed to the mixed results of previous research. Future research should address the lack of appropriate methodology and include larger sample sizes in evaluating the effects of DAP on children's academic achievement. Furthermore, employing a multilevel design would also allow for the examination of other variables that may differentiate the effects of DAP, such as school or state level characteristics.

This study was limited to examining reading achievement as the only outcome variable. Children's performance differs across various subject areas. For instance, whereas girls outperform boys in reading and language arts, boys are superior in mathematics, in general. Therefore, future research should sue the ECLS-K, or other large scale data sets, to examine the effects of DAP on other outcome measure such as mathematics, science, and social skills.

Another area of interest would be the longitudinal examination of DAP from kindergarten through the early elementary years. DAP is a classroom level variable and children usually change their schools or teachers when they move from kindergarten to the elementary school. Therefore, a child's DAP score would be different in the first grade than what it was in kindergarten. This more complex structure of nestedness requires employing a cross-classified random effects model (Raudenbusch & Bryk, 2002). Future research efforts should consider the cross-classification of DAP between grades.

Another possibility for future research concerns the variables that were selected in this study. There could be other variables, not included in this study, that might have mediating effects on DAP. Teacher quality, for instance, has been found to be associated with teachers' utilization of developmentally appropriate practices (Rathbun, Walston, & Hausken, 2000). It has also been documented that professional development activities can have an effect on teachers' use of instructional strategies in the classroom (NCES, 2000). Therefore, future studies should include some teacher characteristics, such as level of education, experience, and teacher certification, in relation to the effects of DAP on child outcomes.

Conclusion

This study attempted to investigate the effects of developmentally appropriate classroom practices on children's reading achievement from kindergarten through third grade. Considering the limitations of previous research methodology, an appropriate multilevel analysis approach was employed. Furthermore, the differential effects of DAP as a function of children's age, gender, family SES, ethnicity, and school sector was examined. Results indicated a consistent pattern in assessing the effects of DAP. Whereas DAP did not have any significant effect on a standardized direct measure of reading, there were significant and positive effects detected at all three grades on an indirect non-standardized measure of reading. The findings contribute to the existing literature in understanding the relationship between DAP and student outcomes by differentiating the effects of two measure that are different in nature. Therefore, decisions about the success of developmentally appropriate practices must consider multiple indicators of student achievement.

Evidence for differential effects of DAP was not enough to conclude that DAP was better for certain groups of children. In particular, there were no systematic patterns of racial differences in children's reading development as a function of DAP. Future research should consider the complex structure of DAP and examine the longitudinal effects of DAP with a cross-classified random effects model.

On the whole, this study provides valuable information to the ongoing debate on the subject of the developmentally appropriate practices in the literature as it utilized a nationally representative large-scale sample. Evidence for the differing effects of DAP on children's reading achievement as measured by two philosophically different measures was revealed.

Since the publication of the NAEYC's position statement (Bredekamp, 1987), developmentally appropriate practice has been debated in the field of early childhood education. The debate is far from over since a much needed growing body of research is still being conducted. However, the debate cannot shadow the fact that the 1987 position statement made a significant contribution in creating an opportunity for increased conversation within and outside the field toward improvement in instructional practices and children's academic achievement. NAEYC continues to answer criticism and misunderstandings or misinterpretations encountered regularly in practice by publishing new position statements and articles in various journals (e.g., Young Children). Bredekamp (1998) addresses major issues of practice that "a revised DAP and our field need to debate in the future: (1) curriculum and assessment; (2) the role of the teacher; (3) the role of culture in development; (4) attention to the individual child; and (5) relationships with families" (p.182). The need for further research, particularly on the longitudinal effects of DAP, has been addressed by many scholars. Future research regarding the psychometric limitations of previous studies would contribute to broaden the understanding of the effects of DAP on young children’s learning.

APPENDIX

CONFIRMATORY FACTOR ANALYSIS OF THE HYPOTHESIZED DAP MODEL

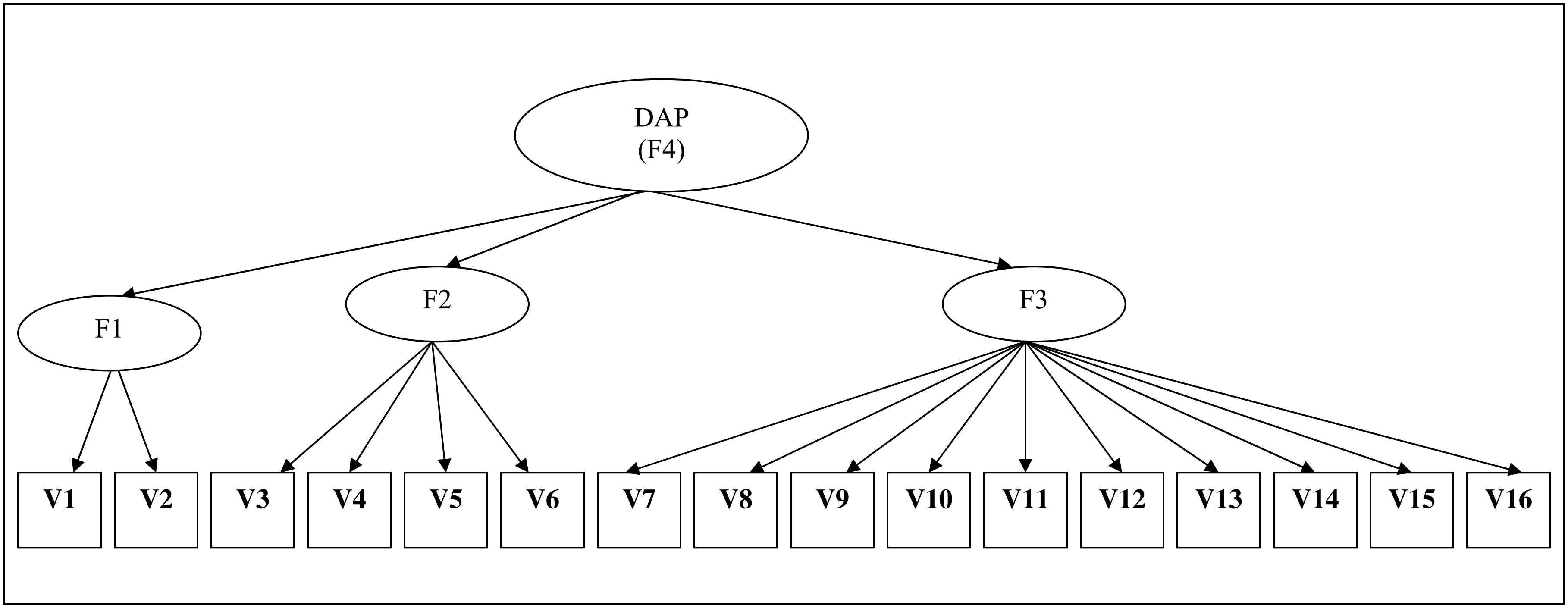

Figure 1. DAP model at the kindergarten level

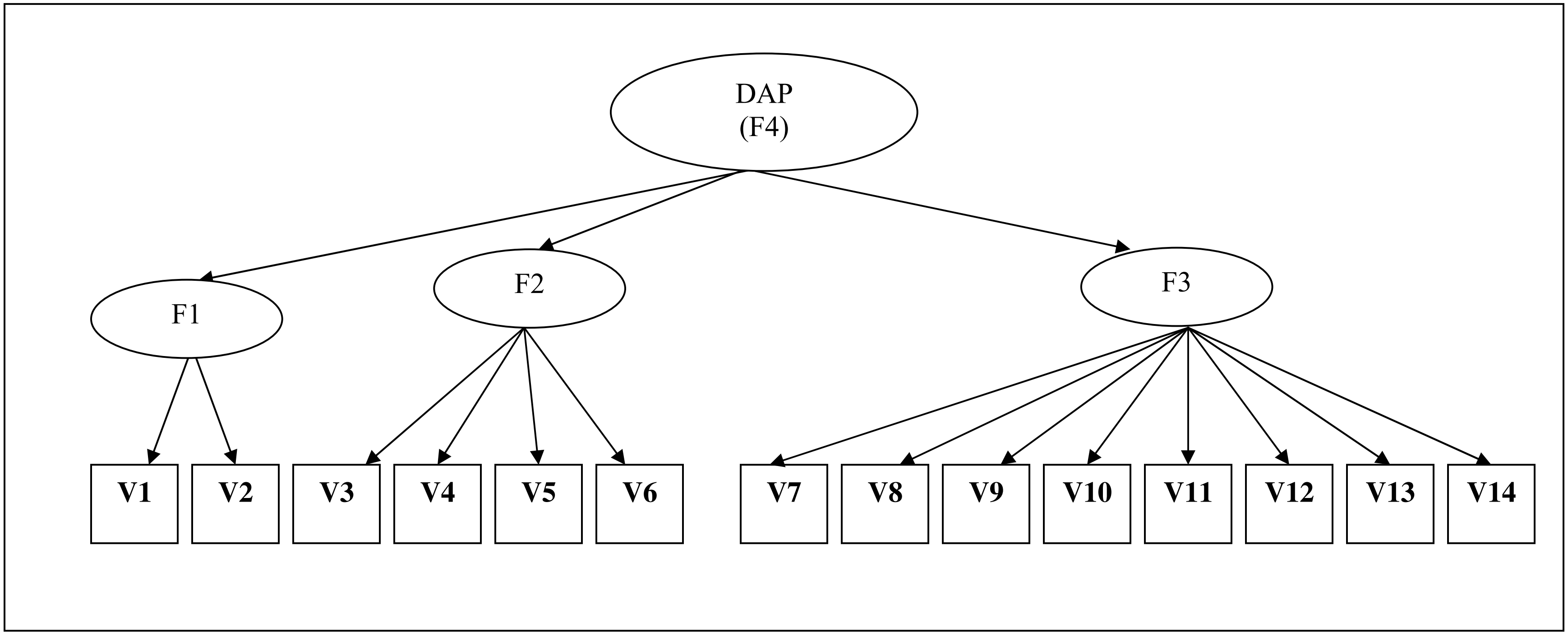

Figure 2. DAP model at the first and third grade levels.

Table 48

Fit Indices for the Three-Factor Hypothesized Model of the Developmentally Appropriate Practices Measure

Model	χ^2	*Df*	CFI	TLI	RMSEA
Kindergarten	2508.61	73	.895	.922	.044
First Grade	1908.20	52	.901	.910	.055
Third Grade	1531.32	59	.881	.891	.053

Note. χ2=Chi-square statistic; df=degrees of freedom for test of model fit; CFI= Comparative Fit Index; TLI= Tucker Lewis Index; RMSEA=Root Mean Square Error of Approximation.

Table 49

Standardized Factor Loadings and Reliabilities for the DAP model at the Kindergarten Level

Factor	Item	Standardized Loadings (λ)	Reliability (R^2)
F1	V1	.683	.466
	V2	.746	.557
F2	V3	.545	.297
	V4	.410	.168
	V5	.655	.429
	V6	.438	.192
F3	V7	.554	.307
	V8	.757	.574
	V9	.452	.204
	V10	.733	.537
	V11	.800	.640
	V12	.566	.321
	V13	.445	.198
	V14	.614	.377
	V15	.669	.448
	V16	.782	.611
F4	F1	.491	.241
	F2	.751	.565
	F3	.846	.716

Table 50
Standardized Factor Loadings and Reliabilities for the DAP model at the First Grade

Factor	Item	Standardized Loadings (λ)	Reliability (R^2)
F1	V1	.616	.380
	V2	.710	.504
F2	V3	.443	.196
	V4	.420	.177
	V5	.641	.412
	V6	.476	.226
F3	V7	.859	.738
	V8	.561	.315
	V9	.730	.533
	V10	.645	.416
	V11	.482	.232
	V12	.608	.370
	V13	.658	.432
	V14	.726	.527
F4	F1	.518	.268
	F2	.662	.438
	F3	.995	.991

Table 51.
Standardized Factor Loadings and Reliabilities for the DAP model at the Third Grade

Factor	Item	Standardized Loadings (λ)	Reliability (R^2)
F1	V1	.549	.302
	V2	.771	.595
F2	V3	.414	.172
	V4	.411	.169
	V5	.510	.260
	V6	.428	.183
F3	V7	.789	.623
	V8	.482	.232
	V9	.684	.468
	V10	.722	.521
	V11	.428	.183
	V12	.570	.325
	V13	.721	.519
	V14	.738	.544
F4	F1	.536	.287
	F2	.677	.458
	F3	.682	.465

REFERENCES

Abbott-Shim, M., & Sibley, A. (1992). *Assessment profile for early childhood programs: Research version*. Atlanta, GA: Quality Assist.

Adams, M. J. (2001). Alphabetic anxiety and explicit, systematic phonic instruction. In S. B. Neuman & D. K. Dickinson (Eds.), *Handbook of early literacy development* (pp. 66–80). New York: Guilford.

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT.

Adams, M. J., & Bruck, M. (1995). Resolving the great debate. *American Educator, 19*, 7-20.

Allington, R.L. (l984). Content coverage and contextual reading in reading groups. *Journal of Reading Behavior, 16*, 85-96.

Anastasiow, N. (1986). The research base for early intervention. *Journal for the Division of Early Childhood, 10*, 99-105.

Blachman, B. A. (2000). Phonological awareness. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3.) (pp. 251-284). Mahwah, NJ: Lawrence Erlbaum.

Bredekamp, S. (1987). *Developmentally appropriate practice in early childhood programs serving children from birth through age eight*. Washington, DC: NAEYC.

Bredekamp, S.(1998). Defining standards for practice: The continuing debate. In C. Seefeldt & A. Galper. *Continuing issues in early childhood education*. Upper Saddle River, NJ: Merril.

Bredekamp, S., & Copple, C. (Eds.)(1997). *Developmentally Appropriate Practice in Early Childhood Programs* (revised edition). Washington, DC: NAEYC.

Bredekamp, S, Knuth, R.A., Kunesh, L.G., & Shulman, D.D. (2004). *What does research say about early childhood education?* Retrieved June 4th, 2004 from http://www.ncrel.org/sdrs/areas/stw_esy/5erly_ch.htm.

Bredekamp, S., & Rosegrant, T. (1992). Reaching potentials: Introduction. In S. Bredekamp & T. Rosegrant (Eds.), *Reaching potentials: Appropriate curriculum and assessment for young children*. Washington, DC: NAEYC

.

Bredekamp, S., & Rosegrant, T. (Eds.). (1995). *Reaching potentials: Transforming early childhood curriculum and assessment*. Vol. 2. Washington, DC: NAEYC.

Brown, A. L., Palincsar, A. S., & Purcell, L. (l986). Poor readers: Teach, don't label. In U. Neisser (Ed.) *The School achievement of Minority Children: New Perspectives.* (pp. 105-143) NJ: Lawrence Erlbaum.

Burts, D.C., Hart, C.H., Charlesworth, R., DeWolf, D. M., Ray, J., Manuel, K., & Fleege, P.O. (1993). Developmental appropriateness of kindergarten programs and academic utcomes in first grade. *Journal of Research in Childhood Education, 8*, 23-31.

Burts, D.C., Hart, C.H., Charlesworth, R., Fleege, P.O., Mosley, J., & Thomasson, R.H. (1992). Observed activities and stress behaviors of children in developmentally appropriate and inappropriate kindergarten classrooms. *Early Childhood Research Quarterly, 7*, 297–318.

Burts, D.C, Hart, C.H, & Kirk, L. (1990). A comparison of frequencies of stress behavior observed in kindergarten children in classrooms with developmentally appropriate versus developmentally inappropriate instructional practices. *Early Childhood Research Quarterly, 5*, 407-423.

Charlesworth, R., Hart, D., Burts, D., & Hernandez, S. (1991). Kindergarten teachers' beliefs and practices. *Early Development and Care, 70*, 17-35.

Clay, M. (1991). *Becoming literate: The construction of inner control.* Portsmouth, NH: Heinemann.

Crawford, P. (1995). Early literacy: Emerging Perspectives. *Journal of Research in Childhood Education, 10*(1), 71- 86.

Curry, N.E., & Johnson, C.N. (1990). *Beyond self-esteem: Developing a genuine sense of human value*. Washington, DC: National Education Association of the United States.

DeVries, R. (1997). Piaget's social theory. *Educational Researcher, 26*, 4-17.

Dickinson, D.K. & Tabors, P. O. (2001). *Beginning Literacy with Language*.Baltimore, MD: Paul Brookes.

Duncan, S. E., & DeAvila, E. A. (1998). *Pre-Language Assessment Scale 2000*. Monterey, CA: CTB/McGraw-Hill.

Dunn, L., Beach, S.A., & Kontos, S. (1994). Quality of the literacy environment in day care and children's development. *Journal of Research in Childhood Education, 9*, 24-34.

Ehri, L. C. (1992). Reconceptualizing the development of sight word reading and its relationship to recoding. In P. Gough, L. Ehri, & R. Treiman (Eds.), *Reading Acquisition,* (pp. 107-143). Hillsdale, NJ: Erlbaum.

Elkind, D.(1981). *The hurried child: Growing up too fast too soon*. Reading, MA: Addison-Wesley.

Elkind, D. (1986). Formal education and early childhood education: An essential difference. *Phi Delta Kappan, 68*, 631-636.

Elkind, D.(1987). *Miseducation: Preschooler at risk*. New York: Knopf.

Elkind, D. (1990). Academic pressures - too much, too soon: The demise of play. In E.Klugman & S. Smilansky (Eds). *Children's play and learning: Perspectives and policy implications* (pp. 3-17). New York: Teachers College.

Erikson, E. (1963). *Childhood and society* (2nd ed.). New York: Norton.

Francis, D. J., Shaywitz, S. E., Stuebing, K. K., Shaywitz, B. A., & Fletcher, J. M. (1996). Developmental lag versus deficit models of reading disability: A longitudinal, individual growth curves study. *Journal of Educational Psychology, 88*, 3-17.

Ferreiro, E., & Teberosky, A. (1982). *Literacy before schooling*. Portsmouth, NH: Heinemann.

Flavell, J.H. (1982). *On cognitive development*. Child Development, 53, 1-10.

Fowell, N., & Lawton, J. (1993). Beyond polar descriptions of developmentally appropriate practice: A reply to Bredekamp. *Early Childhood Research Quarterly,8*, 121-124.

Gardner, J. W. (1990). *On leadership*. New York: Free.

Gelman, R., & Baillargeon, R. (1983). A review of some Piagetian concepts. In P. Mussen (Ed.), *Handbook of child psychology*: *Vol. 3. Cognitive development* (pp. 167-230). New York: Wiley.

Getswicki, C. (1999). *Developmentally appropriate practice: Curriculum and development in early childhood education*. Albany, NY: Delmar.

Goffin, S.G., & Wilson, C.S. (2001). *Curriculum models and early childhood education: Appraising the relationship*. Upper Saddle River, NJ: Merrill Prentice Hall.

Goodman, K. (1986). *What's whole in whole language.* Portsmouth, NH: Heinemann.

Gottlieb, M. (1997). *A Developmentally Appropriate Practice Template administration and technical manual*. Des Plaines, IL: Illinois Resource Center.

Graue, M. E. (1992). Social interpretations of readiness for kindergarten. *Early Childhood Research Quarterly, 7*, 225-243.

Gullo, D.F.(1992). *Developmentally appropriate teaching in early childhood*. Washington, DC: National Education Association of the United States.

Hanson, R. A., & Ferrell, D. (1995). The long-term effects on high school seniors of learning to read in kindergarten. *Reading Research Quarterly, 30*, 908-933.

Harms, T., Clifford, R., & Cryer, D. (1998). *Early Childhood Environment Rating Scale* (Revised Edition). New York: Teacher's College.

Hart, C., Burts, D., & Charlesworth, R. (1997). Integrated curriculum and developmentally appropriate practice birth to age eight. Albany, NY: State University of New York.

Hart, C.H., Burts, D.C., Durland, M.A., Charlesworth, R., DeWulf, M., & Fleege, P.O. (1998). Stress behaviors and activity type participation of preschoolers in more and less developmentally appropriate classrooms: SES and sex differences. *Journal of Research in Childhood Education, 12,* 176-196.

Hemmeter, M. L., Maxwell, K.L., Ault, M.J., Schuster, J.W. (2001). *Assessment of practices in early elementary classrooms (APEEC)*. New York: Teacher's College.

Hirsh-Pasek, K., Hyson, M., & Rescorla, L.(1990). Academic environments in preschool: Do they pressure or challenge young children? *Early Education and Development, 1*, 401-423.

Hoge, R. D., & Butcher, R. (1984). Analysis of teacher judgments of pupil achievement levels. *Journal of Educational Psychology, 76*, 777–781.

Hoge, R. D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of the literature. *Review of Educational Research, 59*, 297–313.

Hoge, R. D. (1983). Psychometric properties of teacher-judgment measures of pupil aptitudes, classroom behaviors, and achievement levels. *Journal of Special Education, 17*, 401–429.

Hopkins, K.D., George, C.A., & Williams, D.D. (1985). The concurrent validity of standardized achievement tests by content area using teachers' ratings as criteria. *Journal of Educational Measurement, 22*, 177-182.

Hox, J.J. (1995). *Applied Multilevel Analysis*. Amsterdam: TT.

Hox, J.J. (1998). Multilevel modeling: when and why. In I. Balderjahn, R. Mathar, & M. Schader (Eds.), *Classification, data analysis, and data highways* (pp. 147-154). New York: Springer Verlag.

Hsue, Y., & Aldridge, J. (1995). Developmentally appropriate and traditional Taiwanese culture. *Journal of Instructional Psychology, 22*, 320-323.

Hu, L., & Bentler, P.M. (1999). Cut-off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55.

Huffman, L.R., & Speer, P.W. (2000). Academic performance among at-risk children: The role of developmentally appropriate practices. *Early Childhood Research Quarterly, 15*, 167-184.

Hyson, M.C., Hirsh-Pasek, K., & Rescorla, L.(1990). The classroom practices inventory: An observation instrument based on NAEYC's guidelines for developmentally appropriate practices for 4- and 5-year-old children. *Early Childhood Research Quarterly, 5*, 475-494.

International Reading Association. (1986). Literacy development and pre-first grade: A joint statement of concerns about the present practices in pre-first grade reading instruction and recommendations for improvement, *Childhood Education, 63*, 110-111.

International Reading Association & National Association for the Education of Young Children. (1998). Learning to read and write: Developmentally appropriate practices for young children. *Young Children*, *53*(4), 3–46.

Isenberg, J.P., & Jalongo, M.B.(1997). *Major trends and issues in early childhood education: Challenges, controversies, and insights*. New York: Teacher's College.

Jambunathan, S., Burts, D.C., & Pierce, S.H. (1999). Developmentally appropriate practices as predictors of self-competence among preschoolers. *Journal of Research in childhood Education, 13*, 167 174.

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology, 80,* 437-447.

Jipson, J.A. (1991) Developmentally Appropriate Practice: Culture, curriculum, connections. *Early Education and Development, 2*, 120–136.

Johnson, E., & Johnson, K. M.(1992). Clarifying the developmental perspective in response to Carta, Schwartz, Atwater, and McConnell. *Topics in Early Childhood Special Education, 12*, 439-457.

Jones, I.., & Gullo, D. (1999). Differential social and academic effects of developmentally appropriate practices and beliefs. *Journal of Research in Childhood Education, 14*, 26–35.

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology, 80,* 437-447.

Kamii, C. (1985). Leading primary education toward excellence: Beyond worksheets and drill. *Young Children, 40*, 3-9.

Kamii, C. (1994). *Young children continue to reinvent arithmetic: 3rd grade*. New York: Teachers College.

Kamii, C., & DeVries, R. (1993). *Physical knowledge in preschool education: Implications of Piaget's theory*. Washington, DC: National Association for the Education of Young Children.

Kamii, C., Manning, M., & Manning, G. (Eds.). (1991). *Early literacy: A constructivist foundation for whole language*. Washington, DC: NEA.

Kenny, D. T., & Chekaluk, E. (1993). Early reading performance: A comparison of teacher based and test-based assessments. *Journal of Learning Disabilities, 26*, 227–236.

Kessler, S. (1991). Alternative perspectives on early childhood education. *Early Childhood Research Quarterly, 6*, 183-197.

Kostelnik, M.J., Soderman, A.K., & Whiren, A.P. (2004). *Developmentally appropriate curriculum: Best practices in early childhood education.* Upper Saddle River, NJ: Pearson.

Krogh, S.L. (1997). How children develop and why it matters. In C. Hart, D. Burts, & R. Charlesworth (Eds.), *Integrated curriculum and developmentally appropriate practice birth to age eight* (pp. 29-48). Albany, NY: State University of New York.

Lloyd-Jones, R., & Lunsford, A.A. (Eds.). (1988). *The English Coalition Conference: Democracy through language*. Urbana, IL: National Council of Teachers of English.

Love, J., Logue, M., Trudeau, J.V., & Thayer, K.(1992). *Transitions to kindergarten in American schools*. Washington, D.C.: U.S. Department of Education.

Lubeck, S. (1994). The politics of developmentally appropriate practice: Exploring issues of culture, class and curriculum. In B. L. Mallory & R. S. New (Eds.), *Diversity and developmentally appropriate practice* (pp. 17-43). New York: Teacher's College.

Ludlow, B.L., & Berkeley, T.R. (1994). Expanding the perceptions of developmentally appropriate practice: Changing theoretical perspectives. In B. L. Mallory & R. S. New (Eds.), *Diversity and developmentally appropriate practices: Challenges of early childhood education* (pp.107-118). New York: Teachers College.

Lyon, G. R. (1999). *The NICHD research program in reading development, reading disorders and reading instruction*. Washington, DC: National Center for Learning Disabilities.

Mallory, B. L., & New, R. S. (Eds.). (1994). *Diversity and developmentally appropriate practice: Challenges for early childhood education*. New York: Teacher's College.

Mantzicopoulos, P.Y., Neuhart-Pritchett, S., & Moreloch, J.B. (1994). *Academic competence, social skills, and behavior among disadvantaged children in developmentally appropriate and inappropriate classrooms*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans.

Marcon, R.A. (1992). Differential effects of three preschool models on inner-city 4-year-olds. *Early Childhood Research Quarterly, 7*, 517-530.

Marcon, R. A. (1993). Socioemotional versus academic emphasis: Impact on kindergartners' development and achievement. *Early Child Development and Care, 96*, 81–91.

Marcon, R. A. (1994). Doing the right thing for children: Linking research and policy reform in the District of Columbia public schools. *Young Children, 50*, 8-20.

Marcon, R. A. (1999). Differential impact of preschool models on development and early learning of inner-city children: A three cohort study. *Developmental Psychology, 35*, 358-375.

Meisels, S.J., Bickel, D.D.P., Nicholson, J., Xue, Y., & Atkins-Burnett, S. (2001). Trusting teachers' judgments: A validity study of a curriculum-embedded performance assessment in kindergarten to grade 3. *American Educational Research Journal, 38*, 73-95.

Miller, L.B.(1996). *The developmentally appropriate inclusive classroom in early education*. Albany, NY: Delmar.

Miller, L., & Bizzell, R.P. (1983). Long-term effects of four preschool programs: Sixth, seventh, and eighth grades. *Child Development, 54*, 727-741.

Morrow, L.M. 1990. Preparing the classroom environment to promote literacy during play. *Early Childhood Research Quarterly 5,* 537-54.

Nagy,W.E., Herman, P.A., & Anderson, R.C. (1985). Learning words from context. *Reading Research Quarterly, 20*, 233-253.

National Association for the Education of Young Children. (1988). NAEYC position statement on standardized testing of young children 3 through 8 years of age. *Young Children, 43*, 42-47.

National Association of State Boards of Education. (1988). *Right from the start: The report of the NASBE task force on early childhood education*. Alexandria, VA: Author.

National Center for Education Statistics (NCES) (2001). *Early childhood longitudinal study, kindergarten class of 1998–99: Base Year Public-use Data Files User's Manual* (NCES 2001–029). Washington, DC: Author.

National Center for Education Statistics (NCES) (2002). *Early childhood longitudinal study, kindergarten class of 1998–99: First-grade Public-use Data Files User's Manual* (NCES 002–134). Washington, DC: Author.

National Center for Education Statistics (NCES) (2003). *Early childhood longitudinal study, kindergarten class of 1998–99: Third-grade Restricted-use Data Files User's Manual* (NCES 2003–003). Washington, DC: Author.

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

National Reading Panel. (2000). *Teaching children to read: An evidence based assessment of the scientific research literature on reading and its implications for reading instruction.* Bethesda, MD: Author.

Neuman, S.B. & K. Roskos. 1992. Literacy objects as cultural tools: Effects on children's literacy behaviors in play. *Reading Research Quarterly, 27*, 202-25.

Nicholls, J.G., Cobb, P., Wood, T., Yackel,E., & Patashnick, M.(1991). Dimensions of success in mathematics: Individual and classroom differences. *Journal of Mathematics Education, 21*, 109-122.

Novick, R. (1996). *Developmentally appropriate and culturally responsive education: Theory in practice*. Retrieved June 4th, 2004 from http://www.nwrel.org/cfc/publications/DAP2.html.

Oka, E., & Paris, S. (1986). Patterns of motivation and reading skills in underachieving children. In S. Ceci (Ed.), *Handbook of cognitive,social, and neuropsychological aspects of learning disabilities* (Vol. 2). Hillsdale, NJ: Erlbaum.

Perfetti, C. (1992). The representation problem in reading acquisition. In P. Gough, L. Ehri, & R. Treiman (Eds.), *Reading Acquisition.* (pp. 107-143). Hillsdale, NJ: Erlbaum.

Perry, N.E., & Meisels, S.J. (1996). *Teachers' judgments of students' academic performance* (Working Paper No. 96-08, National Center for Education Statistics). Washington, DC: U.S. Department of Education.

Piaget, J. (1926). *The language and thought of the child*. New York: Harcourt.

Piaget, J. (1932). *The moral judgment of the child*. New York: Free Press.

Piaget, J. (1952). *The origins of intelligence in children*. New York: International Universities.

Piaget, J. (1981). *Intelligence and affectivity: Their relationship during child development*. Palo Alto, CA: Annual Reviews.

Piaget, J., & Inhelder, B. (1969). *The psychology of the child*. New York: Basic.

Rathbun, A., West, J., & Germin-Hausken, E. (2004). *From kindergarten through third grade: Children's beginning school experiences* (NCES 2004-007). U.S. Department of Education.

Raudenbush, S.W. (2001). Comparing personal trajectories and drawing causal inferences from longitudinal data. *Annual Review of Psychology, 52*, 501-25.

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

Rawl, R.K., & O'Tuel, F.S. (1981). A comparison of three prereading approaches for kindergarten students. *Reading Improvement, 21*, 205-211.

Reitsma, P. (1983). Printed word learning in beginning readers. *Journal of Experimental Child Psychology, 75*, 321-339.

Rock, D.A., & Pollack, J.M. (2002). *Early childhood longitudinal study-kindergarten class of 1998-99 (ECLS-K), psychometric report for kindergarten through first grade*. Washington, DC: United States Department of Education.

Ruckman, A.Y., Burts, D.C., & Pierce, S.H. (1999). Observed stress behaviors of first-grade children participating in more and less developmentally appropriate activities in a computer-based literacy laboratory. *Journal of Research in Childhood Education, 14*, 36-47.

Salvesen, K. A., & Undheim, J. O. (1994). Screening for learning disabilities. *Journal of Learning Disabilities, 27*, 60–66.

Schweinhart, L.J., Weikart, D.P., & Larner, M.B. (1986). Consequences of three preschool curriculum models through age 15. *Early Childhood Research Quarterly, 1*, 15-45.

Sharpley, C. F., & Edgar, E. (1986). Teachers' ratings vs. standardized tests: An empirical investigation of agreement between two indices of achievement. *Psychology in the Schools, 23,* 106–111.

Share, D. (1999). Phonological recoding and orthographic learning: A direct test of the self-teaching hypothesis. *Journal of Experimental Child Psychology, 72*, 95-129.

Silverstein, A. B., Brownlee, L., Legutki, G., & MacMillan, D. L. (1983). Convergent and discriminant validation of two methods of assessing three academic traits. *Journal of Special Education, 17,* 63–68.

Smith, L. (2002). Piaget's model. In U. Goswami (Ed.), Blackwell handbook of childhood cognitive development (pp.394-411). Malden, MA: Blackwell.

Snow, C. E., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy.

Spodek, B. (1991). *What should we teach kindergarten children? Educationally appropriate kindergarten practices*. Washington, DC: National Education Association.

Stafford, M.E., van Rensburg, P.W., & Greene, A. (2000). *Longitudinal efficacy comparisons of three levels of developmentally appropriate practice for former head start children*. Current Issues in Education. Retrieved on June 4th, 2004 from http://cie.ed.asu.edu/volume3/number3/.

Stanovich, K. E. (1986). Cognitive processes and the reading problems of learning disabled children: Evaluating the assumption of specificity. In Torgesen, J., and Wong, B. (eds.), *Psychological and educational perspectives on learning disabilities* (pp. 87–131). New York: Academic.

Stevenson, H. W., & Newman, R. S. (1986). Long-term prediction of achievement and attitudes in mathematics and reading. *Child Development, 57*, 646–659.

Stipek, D., & Byler, P. (2004). The early childhood classroom observation measure. *Early Childhood Research Quarterly, 19*, 375–397.

Stipek, D.J., Feiler, R., Daniels, D., & Milburn, S. (1995). Effects of different instructional approaches on young children's achievement and motivation. *Child Development, 66*, 209-223.

Stipek, D. J., Feiler, R., Byler, P., Ryan, R., Milburn, S., & Salmon, J. M. (1998). Good beginnings: What difference does the program make in preparing young children for school? *Journal of Applied Developmental Psychology, 19*, 41–66.

Tabachnick, B.G., & Fidel, L.S. (1996). *Using multivariate statistics* (3rd ed.). New York: Harper Collins.

Torgesen, J. K., & Burgess, S. R. (1998). Consistency of reading-related phonological processes throughout early childhood: Evidence from longitudinal-correlational and instructional studies. In J. C. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning reading* (pp. 161-188). Hillsdale, NJ: Lawrence Erlbaum.

Torgeson, J. K., & Davis, C. (1996). Individual difference variables that predict response to training in phonological awareness. *Journal of Experimental Psychology, 63*, 1–21.

Torgesen, J.K., Rashotte, C.A., & Alexander, A. (2001). Principles of fluency instruction in reading: Relationships with established empirical outcomes. In M. Wolf (Ed.), *Dyslexia, fluency, and the brain* (pp. 333-355). Parkton, MD: York
.

Torgesen, J. K., Wagner, R., & Rashotte, C. A. (1994). Longitudinal studies of phonological processing and reading. *Journal of Learning Disabilities, 27*, 276-286.

Van Horn, M.L., & Ramey, S.L. (2003). The effects of Developmentally Appropriate Practices on academic outcomes among former Head Start students and classmates from first through third grades. *American Educational Research Journal, 40*, 961–90.

Van Horn, M.L, Karlin, E.O., Ramey, S.L., Aldridge, J., & Snyder, S.W. (2005). Effects of developmentally appropriate practices on children's development: A review of research and discussion of methodological and analytic issues. *The Elementary School Journal, 105*, 325-351.

Vygotsky, L.S. (1978). *Mind in society*. Cambridge, MA: Harvard University.

Walsh, D.J. (1991). Extending the discourse on developmental appropriateness. *Earl Education and Development, 2*, 109-119.

West, J., Denton, K., & Germino-Hausken, E. (2000) *America's kindergartners*. Washington, DC: U.S. Department of Education, Report (NCES2000-070).

Wolfgang, C.H., & Wolfgang, M.E. (1999). *School for young children: Developmentally appropriate practices*. Needham Heights, MA: Allyn & Bacon.

BIOGRAPHICAL SKETCH

Alper Tolga Kumtepe was born on July 31th 1972 in Bilecik, Turkey. He received his undergraduate degree in Educational Communications in 1995 at the Anadolu University, Eskisehir, Turkey. After working as a research and teaching assistant at the same program for two years, he was awarded an overseas graduate scholarship by the Turkish Ministry of Education. He earned his M.S. degree in 2000 and Ph.D. degree in 2005 in the program of Early Childhood Education at the Florida State University.

During his final year in the doctoral program, he worked as a researcher at the Florida Center for Reading Research. His research interests include child development, teacher education, quantitative research methods, and the education of gifted young children.