EXAMINING THE INFLUENCE OF EARLY HOME LITERACY ENVIRONMENT AND SOCIOEMOTIONAL COMPETENCE DEVELOPMENT ON READING ACHIEVEMENT

A Dissertation

by

XUEJUN JI

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	R. Malatesha Joshi
Committee Members,	Emily Cantrell
	Wen Luo
	Kausalai Wijekumar
Head of Department,	Michael A. de Miranda

August 2017

Major Subject: Curriculum and Instruction

ABSTRACT

The purpose of this study is to (1) identify the underlying types of home literacy environment (HLE) and their effects on early reading achievement, (2) examine the types of socio-emotional competence (SEC) and their long-term effects on later reading achievement, and (3) investigate the contribution of the HLE and SEC to identify the at-risk readers. Latent Variable Modeling and Decision Tree were used for data analysis.

The sample in this study consisted of 13,367 early graders extracted from the Early Childhood Longitudinal Study Kindergarten dataset (ECLS-K) 2010-2011.
The reading achievement variables are the IRT-scaled reading scores at the kindergarten fall and first grade spring. HLE measures include 13 survey items regarding direct and indirect literacy interaction between children and parents. SEC measures are the social skill items based on parent-reported and teacher-reported scales for both grade levels.

Two main types of home literacy environments were identified: pro-reading and contra-reading families. Further analysis indicated that over 70% of the sampled families provided supportive reading environments. Children from the pro-reading families outperformed their contra-reading families in reading achievement. Furthermore, results indicated that the racial groups and socioeconomic status (SES) also differentiate the degree of parental involvement into home literacy activities. Next, students from higher SES families tended to be more likely to experience rich home literacy, and thus obtained early advantage in reading achievement. Meanwhile, these children were more likely to stay in and move to the positive SEC behavior state in comparison with their

counterparts from the low SES families, and the negative SEC behavior state was also associated with low reading achievement. Finally, the at-risk reader profile was identified in the Decision Tree Model.

In conclusion, this study examined the effects of the ecological and psychological factors on early reading achievement, and attempted to build a predictive model to identify and profile the at-risk readers. When the students are observed to have the profiled behavior, they might be further diagnosed by a reading specialist and a school psychologist for verification. The teacher can give special attention, extra instruction, and immediate intervention to that particular group.

DEDICATION

I dedicate my dissertation work to family and friends. A special gratitude to my beloved parents, Hongmei (Theresa) Xie, and Youshen Ji whose words of encouragement and exhortation echo in my ears.

I also dedicate this dissertation to my many friends who have supported me throughout the course of pursuing my doctorate degree at Texas A&M University, especially Luxi Feng and Shuqiong (Linda) Lin for accompanying me and cheering me up during the intensive academic training. Both of you have been my best cheerleaders.

Finally, I dedicate this work and give special thanks to Dr. Dmitri V. Voronine for facilitating my academic success throughout the entire process.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Joshi, and my committee members, Dr. Luo, Dr. Cantrell, and Dr. Wijekumar, for their guidance and support throughout the course of this research.

Thanks also go to colleagues and the department faculty and staff for making my time at Texas A&M University a great experience.

Finally, special thanks to my mother and father for their encouragement and support and to my "best brothers forever" – Xichen (Alex) Wang, Dmitri V. Voronine, and Dongdong (Tony) Ge for being the friends that I always wanted, needed and appreciated.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Professor Joshi and Professors Wijekumar and Cantrell of the Department of Teaching, Learning and Culture and Professor Luo of the Department of Educational Psychology.

All work for the dissertation was completed independently by the student.

Funding Sources

Graduate study was supported by the CEHD Strategic Research Fellowship from College of Education and Human Development, Texas A&M University

NOMENCLATURE

AIC	Akaike Information Criterion
AWE	Approximate Weight of Evidence
BIC	Bayesian Information Criterion
CAIC	Consistent Akaike's information Criteria
CIs	Confidence Intervals
df	degree of freedom
HLE	Home Literacy Environment
LL	Lowe Limit
MOE	Margin of error
OR	Odds Ratio
s.e.	standard error
SEC	Socioemotional Competence
SES	Socioeconomic Status
UL	Upper Limit

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Page

Page

Page

CHAPTER I

INTRODUCTION

The home literacy environment (HLE) and socioemotional competence (SEC) are the essential factors that affect students' early reading development. Researchers have paid growing attention to the conceptualization of HLE and SEC, and they have been defined in various ways. For example, HLE is predominantly conceptualized by a single measure—the frequency of shared reading between parents and children (Bus, van IJzendoorn, & Pellegrini., 1995). However, one single measure cannot capture the multidimensional nature of HLE. Accordingly, other researchers define HLE as being multifaceted. Specifically, in addition to the shared reading, the measures of HLE also include, for example, the number of picture books at home, the frequency of child's request for reading at home, the frequency of visits to libraries, and parents' reading to children at home (Griffin & Morrion, 1997; Scarborough, Dorich, & Hager, 1991). Likewise, SEC does not have a consensus of definition either. Researchers from education, sociology, economics and psychology define this construct with different terms, even though some scholars adopted the same terms but the construct emphasized different aspects. (Blair & Razza, 2007; Diprete & Jennings, 2012; Heckman & Rubenstein, 2001; Morgan, Farkas, & Wu, 2001).

Admittedly, the inclusions of the HLE and SEC measures are by no means exhaustive, but the underlying dimensionality could be identified. However, few of the research studies have successfully identified the multidimensional nature of HLE and

SEC profiles, since the researchers have employed univariate variable-centered analysis. Hence, the variations between each subtype of HLE and SEC have been rarely examined, let alone their associations with those demographic factors and reading skills.

In addition to the conceptualization and measurement issues of the HLE and SEC profiles, much extant literature only examined the predictive power of the early SEC profile on later reading skills. Few studies have examined its transition over time, and the effects of the transition response patterns on later reading achievement. Finally, there is also a lack of investigation into the complex interactions among HLE, SEC, and other demographic factors on reading achievement, of which the results are conducive to identifying warning system for at-risk readers.

To address the aforementioned issues, this dissertation project uses Componential Model of Reading (CMR; Joshi & Aaron, 2000, 2012; Joshi, Tao, Aaron, & Quiroz, 2012) and Ecological System Theory (Bronfenbrenner, 1986, 1994) to shape the scope and direction of the study. Finite Mixture Model and Decision Tree are used to examine the research goals in the three studies as follows:

1) To identify the early home literacy environment (HLE) profile and its effects on early reading achievement.

2) To identify the student's socioemotional competence (SEC) profile and its impacts on reading achievement.

3) To examine the combined effects of HLE and SEC on later reading achievement.

Research Questions

This dissertation research project includes three interconnected studies. Specifically, these studies address three issues regarding how the factors in ecological domain and psychological/socioemotional domain affect the reading achievement: 1) the latent profile of HLE and its association with the reading achievement; 2) the underlying types of socioemotional competence (SEC) and its stability over grades; and 3) the combined effects of HLE and SEC on reading achievement.

My first study attempts to conceptualize the students' HLE profile and its association with the early reading achievement. The main goal of Study 1 is to conceptualize the home literacy environment by including the constructs: shared reading activities, access to literacy materials, and parental involvement in cognitive stimulation activities. The second goal is to examine the variation of the HLE profile after adding SES and race/ethnicity into the model. The third goal is to investigate the associations of the HLE profile with the early reading achievement.

For Study II, I am going to conduct Latent Class Analysis on socioemotional competence (SEC) measures to identify the underlying types of the students' socioemotional behavior, and examine the variation in classification solution of SEC on covariates including SES, and gender. Moreover, I will also test the stability of the SEC profile over grades, and detect the difference in the reading achievement associated with the transition response patterns.

For the third study, I incorporate the obtained latent discrete variables (i.e., HLE and SEC), and variables of the transition response patterns to investigate the confluence of HLE and SEC and its interaction with other factors (e.g., demographic factors and early reading ability at kindergarten entry) on the later reading achievement. Cumulatively, these studies attempt to answer the research questions detailed in Table 1 and the visual representation of the hypothesized models for Study 1 and 2 are presented in Figures 1 to 3. (Note: The Study 3 is a data-driven analysis using Decision Tree method, so there is no hypothesized model for this study.) Visual representation of the hypothesized models for Study 1 and 2 are presented in Figures 1 to 3 below.

Table 1

The Purpose and Research Questions for Each Study

Purpose of Study	Research Questions
Study 1: To identify early home literacy environment (HLE) profile and its effects on early reading achievement.	1. What are the types of HLE profile of students at the entry of kindergarten (Kindergarten, Fall)? 2. Does the membership of the identified HLE profile vary upon SES (i.e., family income, maternal education) and membership in a racial group? 3. What is the association between HLE and students' early and later achievement after controlling for SES and membership in racial and ethnic groups
Study 2: To identify students' socio-emotional profile and its impacts on reading achievement	4. What is the latent profile of SEC, and its variation in membership on covariates including SES and gender? 5. What is the association between the socioemotional competence (SEC) and students' reading achievement? 6. Is the SEC profile stable over time? 7. What are the long-term effects of SEC on later reading achievement?
Study 3: To examine the combined effects of HLE and SEC on later reading achievement	8. What is the variation in later reading achievement predicted by the profiles of HLE and SEC, and SEC transition response patterns, as well as their interaction with other covariates?

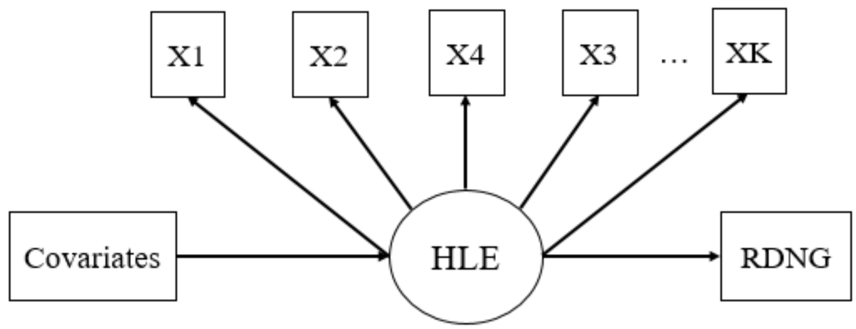


Figure 1 The Hypothesized Model for Study 1

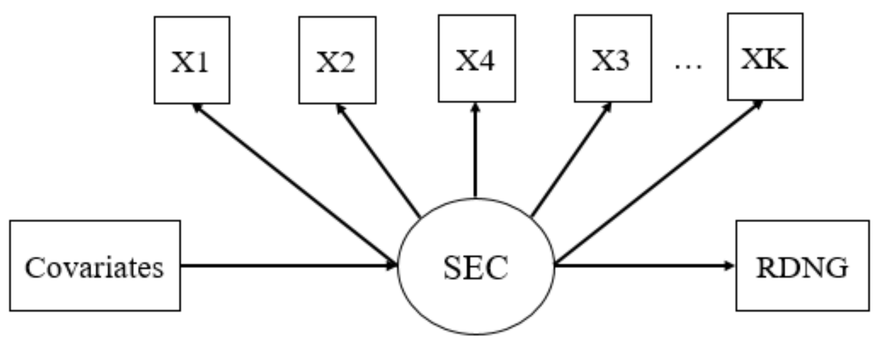


Figure 2 The Hypothesized Model for Study 2

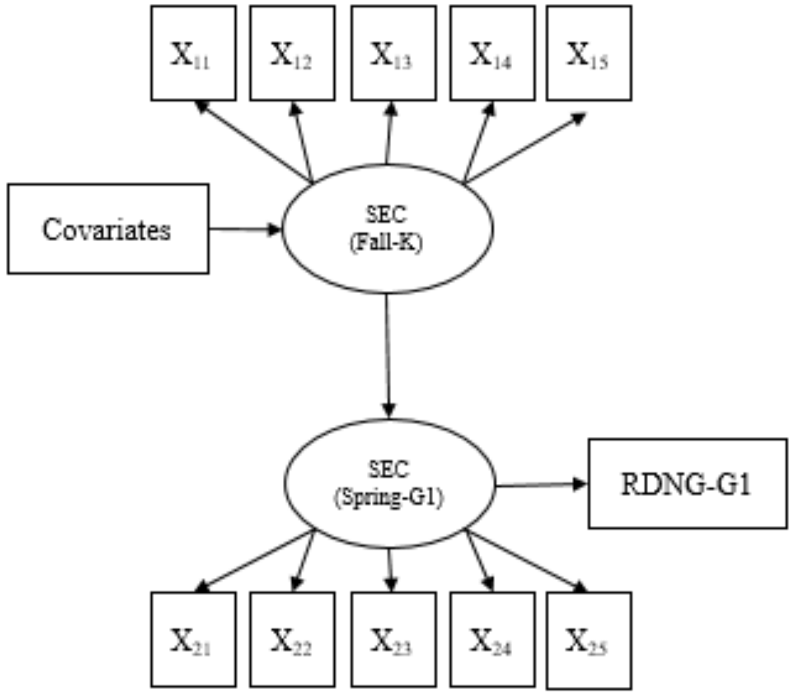


Figure 3 The Hypothesized Latent Transition Model

Limitations

This study has some limitations. First, this study did not incorporate the complex survey sampling weights during the analysis, because the complexity of the model estimation may easily lead to the failed model identification and model convergence. Also, the results of this study cannot be generalized to the entire American K-1 students, but only to the extracted sampled participants. Second, the survey items selected in this research were not culturally adaptive. To be more specific, the HLE items did not include sufficient information regarding indirect home literacy activities, which are the typical literacy features in Asian American families. The limited selection of survey items is due to the design of ECLS-K survey design. Hence, the validity of the selected items may be slightly contaminated. Third, the cut-off value for at-risk readers was fixed at the bottom 10% of the distribution based on the extant literatures. However, different cut-off values might also lead to different classification results. Hence, the obtained results in study 3 should be evaluated in comparison with other conditions.

Significance of the Study

Despite the limitation as mentioned above, this study still successfully identified the underlying multi-faceted structure of home literacy environment (HLE) and socioemotional competence (SEC), and its associations with the later reading achievement. In addition, the variations in HLE and SEC by family backgrounds were also explicitly identified. It is, in fact, one of the very few studies with the focus on the attributes of the individual using person-centered analysis. Moreover, it also filled the gap in extant literacy study literatures exploring the association between the socio-emotional behavior and its over-time transition and reading achievement. Finally, the study built up a predictive system for at-risk readers based on the ecological and psychological domains of the Componential Model of Reading (CMR). The Decision Tree format results are easier for practitioners to understand, and useful for the classroom practice.

CHAPTER II

LITERATURE REVIEW

In the opening of Chapter I, I have attempted to motivate and outline my study of the interconnections among home literacy environment (HLE), socioemotional competence (SEC), and early reading achievement. In this chapter, I review the literatures on two separate domains: theoretical framework and methodology. In terms of the first domain, I first clarify the concepts of theoretical frameworks related to this study: Componential Model of Reading (CMR; Joshi & Aaron, 2000, 2012; Joshi, Tao, Aaron, & Quiroz, 2012) and Ecological System Theory (Bronfenbrenner, 1986, 1994). Next, I define and shape the scope of this study by reviewing the extant literatures relating to the theoretical models. Finally, I evaluate the research so far, and examine the gaps in the body of research. Concerning the second domain, I briefly introduce the Latent Variable Modeling Methods (e.g., Latent Class/Profile/Transition Analysis) and Decision Tree.

Review on Theoretical Framework

Componential model of reading. Simple View of Reading (SVR, Gough & Tunmer, 1986; Hoover & Gough, 1990) is a useful model to examine the associations between emergent literacy skills and reading comprehension (i.e., Decoding × Linguistic Comprehension = Reading Comprehension). This model has been widely examined in English language, and other orthographies. (e.g., Joshi, in press; Joshi, Ji, Breznitz, Amiel, & Yulia, 2015). However, such model only focuses on cognitive skills, and does not include other factors that also contribute to the reading achievement such as home [literacy] environment, motivation, peer influence, and the classroom environment.

Componential Model of Reading (CMR; Joshi & Aaron, 2000; Aaron, Joshi, & Quatroche, 2008; as shown in Figure 4) extends SVR to a three-component model by including ecological and psychological domains in addition to the cognitive domain, which conceptualizes more broad factors that contribute to the reading development. The psychological domain within CMR includes, for example, motivation and interest, teacher's expectation, and gender differences, and the ecological domain contains teacher knowledge, dialect differences, home [literacy] environment, and orthography. In short, the CMR attempts to incorporate various factors in a cohesive, yet comprehensive fashion to model how the three-domain factors influence the acquisition of literacy skills, and, in turn, to allow for differential diagnosis and treatment of reading problems (Joshi & Aaron, 2012).

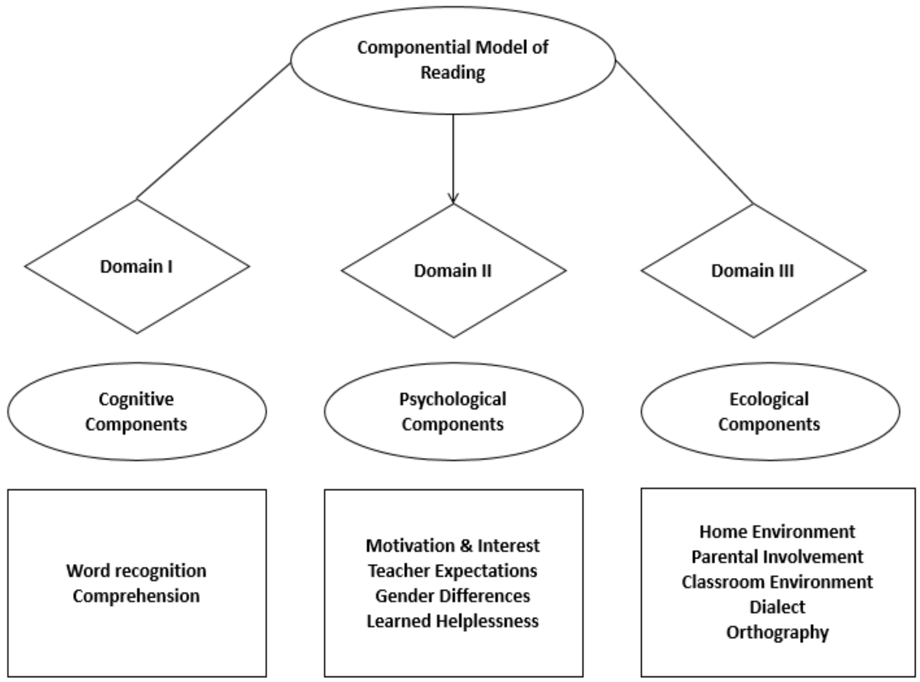


Figure 4 Componential Model of Reading. Adapted from *Becoming a professional reading teacher* (p.11), by P.G. Aaron, R.M. Joshi and D. Quatroche, 2008, Baltimore, MD: Paul H. Brookes Publishing Co. Copyright © 2008 by the Paul H. Brookes Publishing Co. Adapted with permission.

A special issue in the *Journal of Learning Disabilities* (2012) addressed how the factors of cognitive, ecological and psychological domains contribute to reading acquisition and reading difficulties within the framework of the CMR. In that issue, for example, Chiu, McBride-Chang, and Dan (2012) examined the contributions of three domains on the reading performance of 186,725 fourth graders from 38 countries, and found that the ecological domain (i.e., SES, parental attitude, school characteristic, number of books at home, and enjoyment of reading) explained around 90% of variance in reading performance. Hernandez, Folsom, Al Otaiba, Greulich, Thomas-Tate, and Connor (2012) applied the CMR to explore how kindergarten-entry factors predicted first grade reading performance. In contrast, these authors found that the ecological domain factors accounted for roughly 20% of the variance in the first grade reading performance, the psychological and cognitive domains explained about 18% and 16% of the variance, respectively, suggesting that it is important to examine all three domains in the diagnosis and intervention of reading problems.

Still, fewer studies in extant literacy research have examined the contributions of ecological domain and psychological domain to the reading development in comparison with research relating to the cognitive domain. Hence, one goal of this research project is to validate the ecological and psychological domains within CMR.

Ecological system theory. How are children's development and growth affected by the environment around them? The ecological system theory, which was proposed by Bronfenbrenner (1986, 1994) has addressed this question. According to Bronfenbrenner, children's development was influenced by everything around them. The children's behavior, thus, was viewed as a function of interactions with their environment. And this surrounding environment encompasses at least four interacted and nested systems: microsystem, mesosystem, exosystem, and macrosystem (as shown in Figure 5).

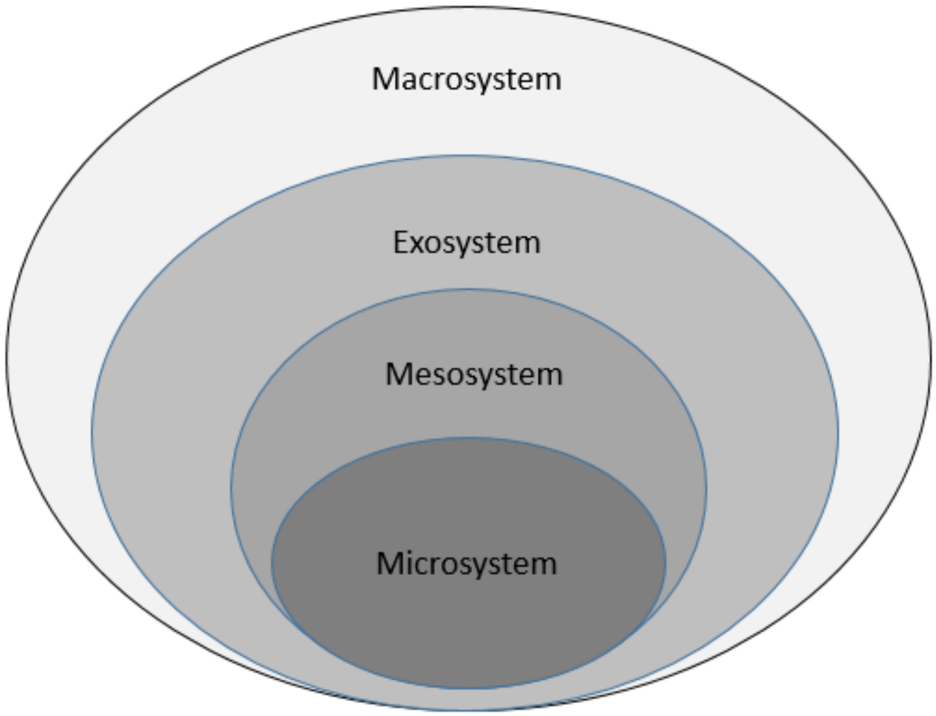


Figure 5 Ecological System Theory

At the microsystem is a child's individual interaction and relation within a particular social setting, for example, home environment and socioemotional behaviors. Applied to the reading acquisition, the home environment is crucial to children's reading development in particular to those children before formal schooling, because the interactions in literacy activities mostly occurred between parents and children at home (McBride, 2015). Additionally, the home socioemotional climate also plays an important role in promoting book reading (Leseman & de Jong, 1998), and reciprocal relation has also been found between literacy development and socioemotional competence. For example, shared reading activities with the contents of social and emotional promotion can facilitate both reading and socioemotional development (Jones, Brown, & Lawrence

Aber, 2011). Thus, the home literacy environment and socioemotional development can be categorized as the factors at the microsystem level.

The mesosystem consists of interacted associations between individual microsystems. For instance, the mesosystem could refer to the parental involvement in children's school activities such as attending the parent/teacher conference, and volunteering in school events, which reflects the interaction between parent-child relation and teacher-child relation (McBride, 2015).

The next higher level is exosystem, which refers to those indirect effects in the environment on children's development. For example, the job promotion of children's parents or school retention policy could be classified as the factors at this level.

The macrosystem encompasses the cultural contexts of individuals. The factors at this level may include race/ethnicity, SES, and religious affiliations. In reference to literacy acquisition, the cultural contexts (e.g., race/ethnicity) have been found to differentiate parental expectations and self-concept towards achievements across ethic groups. For example, Asian students carry very high expectation from their parents for academic success, as a result, bearing a very strong pressure to succeed academically. Asian parents may also attribute academic success to diligence or hard work. In contrast, American (Caucasian) parents are more likely to view the academic performance as the result of inherent ability (McBride, 2015). In addition, the macrosystem level also includes the gender differences. For instance, in a recent study, Gracia (2015) found noticeable gender disparities in socioemotional competence, using Early Childhood Longitudinal Study (ECLS-K: 2011) dataset. Specifically, the study found girls had relative advantage over boys in creativity, closeness to teachers, and behavior control at the starting point of formal school.

The following sections of this chapter focus on how the factors from macrosystem and microsystem extracted from Bronfenbrenner's ecological system theory affect children's reading abilities. At the macro-level, I shall examine how SES and race differentiate the reading ability. Next, I shall investigate the influence of the SES and race/ethnicity on the HLE (micro-level), and their further impacts on the reading achievement.

Macrosystem level: SES and race/ethnicity. Before entering school, U.S. children reading skills vary widely by their SES and race/ ethnicity. Cumulative evidence has shown that children from low SES family and from minority families have substantially lagged behind their higher SES counterparts in the emergent literacy skills. And these children with advantageous literacy skills are often raised by more educated parents. (Hecht, Burgess, Torgesen, Wagner, & Rashotte, 2000; Lonigan, Burgress & Anthony, 2000; Raz & Bryan, 1990, Reardon, 2011). These sociocultural gaps are also consistent with the findings from some nationally representative data studies. Lee and Burkam (2002) analyzed the gap in the early reading skill and its association with a composite SES measure, using data from the first wave (i.e., kindergarten in the fall of 1998) of the ECLS-K: 1998 dataset. They observed that sizeable gaps in reading and math arise at the school entry in the kindergarten. For example, their findings showed that the average score in reading and math in the highest SES group is 60% higher than that of the lowest SES group (However, after controlling for the SES, the race-related gap is not statistically significant). In a recent study, Garcia and Weiss (2015) reported similar patterns. Using the first wave of the ECLS-K: 2011 data, they also explored the initial gaps in the reading skills and found the substantial difference in early literacy skills by replicating the study by Lee and Burkam (2002). Specifically, their results indicated that children from the high SES family have significantly higher reading scores than their peers in the low SES group. The reading skills advantages of middle SES children over the low SES group are about half as large as the advantages of the high SES to the low SES group. Considering the racial groups, substantial gaps in the reading skills exist when comparing Asians and Caucasians with African-Americans and Hispanics. And Asian children also have significant relative advantages over Caucasians. Similar to Lee and Burkham's (2002) findings, after controlling the SES differences, the reading skills gaps shrink between some racial group comparisons: Asian-Caucasian gap, and Caucasian-African-American gap. The diminishing race-based gaps also suggested that the SES classification is highly associated with race/ethnicity groups. With one exception, the only group comparison that showed a highly significant inequality is the Caucasian-Hispanic (ELL) gap.

Microsystem level: Home literacy environment. Since the disparities exist before children receive formal schooling, the gaps must arise from factors out of formal schooling (e.g., home environment), and might continue to influence their reading development through school years (Waldfogel, 2012). The results of the empirical studies support the key role of HLE in explicating the SES and socio-cultural difference in literacy skills, and consistent findings suggest that the frequency of HLE activities (e.g., storybook telling) differ by social classes and social cultural backgrounds. For example, the recent study by Hamilton, Hayiou-Thomas, Hulme, and Snowling (2016) found that children from families with higher SES tend to have more storybook exposure at home. Similar findings have also been documented by Hemmerechts, Agirdag, and Kavadias (2016). Their studies also demonstrated that students with lower SES had less parental involvement in the literacy activities than their peers from more affluent families. One possible explanation could be that "[parents in low SES family] 'lived busy and satisfying lives with very little mediation by print' (Purcell-Gates, 1996, p.425) and thus spent little if any time reading with or exposing their children to print" (Phillips & Lonigan, 2009, p148).

In addition, HLE activities also differ by sociocultural contexts. The Federal Interagency Forum on Child and Family Statistics (2003, as cited in Philips & Lonigan, 2009) reported that 64% Caucasian family read to their children daily, and 48% of African-American and 42% Hispanic parents reported did so. More empirical evidence has also been documented by other researchers (Bradley, Corwyn, McAdoo, & García Coll, 2001; Brooks-Gun & Markman, 2005; Yarosz & Barnett, 2001). For example, in Bradley et al's study, it was found that a higher percentage of African-American and Hispanic children did not have books at home, and lower percentage had 10 or more books. In contrast, Caucasian children are more likely to possess many more books, and other language learning materials and devices. However, racial differences in HLE will shirk after controlling for SES, which means SES has a much larger impact on the HLE. Yet it should be noted that there is still substantial variation of HLE in SES categories and social cultural groups. Even though for lower SES families, with the limited access to literacy resources, many parents still do provide supportive HLE for children. (Drummond & Stipek, 2004; Payne, Whitehurst, & Angell., 1994; Purcell-Gates, 1996). A recent study by Tichnor-Wagner, Garwood, Bratsch-Hines, and Vernon-Feagans (2016) provided additional evidence. The sample in their study was rural families with annual income of less than 20,000 U.S. dollars. However, their study showed that more than 90% of the families reported that parents provided assistance to homework at least twice a week, over 80% reported that family members supported the child in learning to read at least twice a week. Their findings do not support the negative stereotype that low-income parents do not get involved with helping their children with their academic success in school, and also the claim that parents in low income families do not care about the education of the children. However, one should be cautious about concluding that certain racial groups must value education or not by ignoring the variation and individual differences within those groups. For example, one study with Caucasian, Hispanic, African-American families, showed that all parents thought the emergent literacy skills (knowing letters) to be as important as social-behavior skills for kindergarten success (Diamond, Reagan, & Bandyk, 2000). Hence, it is risky to delineate the features of HLE by only considering the SES and racial between-groups differences, without exploring the within-group variation.

Microsystem level: Socioemotional competence development. In addition to the academic skills, parents have concerns about students' socioemotional competence**,** which might substantially contribute to children's early school success. Using psychometric measures, Kieffer, Vukovic and Berry (2013) found that attention shift and inhibitory control have unique direct contribution to reading comprehension after controlling for working memory, processing speed, and phonological awareness. Howse, Lange, Farran, and Boyles (2003) used teacher's rating measures to investigate the relationship between socioemotional skills and reading achievement, and found that children's motivation predicted concurrent reading achievement. Using different teacher's rating measures, Ladd, Kochenderfer, and Coleman (1996) found that students' engagement and independence predicted their academic success and progress in the early school years.

Several longitudinal studies have also revealed that the early socioemotional competence predicted the concurrent reading achievement as well as the later reading progress. For example, Stipek, Newton, and Chudgar (2010) reported that the positive social-emotional competence (e.g., work independently, seek challenges, accept responsibility for a given task, tuned in to what's going on in the classroom) in the kindergarten and first grade can promote literacy achievement and even in the sequent grade. Furthermore, researchers also found the initial gap between the poor socioemotional competence of children and their higher-rated peers was widening as students entered upper grades (McClelland, Acock, & Morrison, 2006).

Even though researchers have examined different aspects of the socio-emotional competence, using different rating measures and various designs (i.e., cross-sectional or longitudinal), these studies have provided convergent and cumulative evidence to show that the early socioemotional competence is associated with children's reading achievement and their progress in the subsequent grades.

In addition, a few of studies also have examined the group differences (e.g., gender, SES and race) in the socioemotional competences. Using the ECLS-K: 1998 dataset, DiPrete and Jennings (2012) found that girls have an advantages over boys in the SEC in kindergarten, and this advantage continues to grow through the first six years in elementary school. Yet another group of scholars (Buchmamn & DiPrete, 2006; Downey &Vogt Yuan, 2005; Farkas, Grobe, Sheehan, & Shuan, 1990) suggested that female adolescent students sustained their relative advantage of SEC to their male counterparts till college. And such gender disparities in SEC also related to the gender gap in the academic achievement and college completion. As for the influence of SES gap, previous studies suggested children from the impoverished families were rated much lower than their peers who have better living condition. (Dearing, McCartney, & Taylor, 2006). In contrast, the study by Howse el al. (2003) did not find the significant difference in rating children's social-emotional behaviors. Finally, race-based gap was also identified in the socioemotional competence development. For example, McClelland, Morrison, and Holmes (2000) found that ethnic minority students tend to be rated lower and most of their parents have lower education, and less paying jobs. Similarly, Connell and Prinz (2002) reported that ethnic minority children had lower-rated behavior regulation. However, consistent with the race-based gap in the literacy skills, after controlling for SES, this gap diminishes and even disappears, due to the high correlation between SES and race.

Gaps in the literatures. However, most of the research predominantly focuses on the associations between variables, and fails to give a clear picture of the HLE and SEC profiles. Thus, the variations between each subgroup were rarely examined, let alone their association with demographic factors and reading skills. Moreover, the multi-dimensional nature of HLE and SEC has been ignored in most studies by using single measure or one single composite score. Even though some studies have used multiple measures but have ignored the multivariate associations by using simple linear regression.

In the remaining section, I am going to synthesize the conceptualization of HLE and SEC from the extant studies to determine the scope of selecting measures for the current study. Next, the rationale of using the person-centered method (e.g., Latent Class Analysis) and Decision Tree will also be briefly explained.

Conceptualization of HLE. The construct of home literacy environment (HLE) is predominantly measured by a single measure—shared reading activity (e.g., read books to children) between parents and children (Bus, IJzendoorn, & Pellegrini, 1995; Philips & Lonigan, 2009). However, a single measure may not reflect the multi-dimensional nature of HLE, as well as the validity and power of using a single variable measuring a complex construct which is highly questionable (Scarborough & Dobrich, 1994). Instead, some scholars conceptualized HLE as being multi-faceted. For example, in addition to the shared reading activity, HLE also includes the dimension of access to the literacy sources. The related measures include the number of picture books at home, the frequency of child's request for book reading at home, and the frequency of library visits (Payne, Whitebust, &Angell, 1994; Scarbourgan, Dorich & Hager, 1991). Moreover, cognitive stimulation activities are also considered to be as another dimension of HLE including child watching educational television (hours per day; Philips & Lonigan, 2009), playing word games, singing songs or telling stories to children (Hemmerchts, Agirdag, & Kavadias, 2016; Wheaton, 2010). The current study adopts the three-dimensional model including shared reading, access to literacy materials, and stimulant activities. The model is grounded by the availability of measures in the ECLS-K:2011 dataset.

Conceptualization of SEC. Even though DiPerna and his colleagues (2007) have identified four types of student behavior as the predictors of students' academic skills: externalizing behavior problem, internalizing behavior problem, interpersonal skills, and approaches to learning, yet there is still a lack of consensus term to define the SEC, which reflects the multidimensional nature of these skills (DiPrete & Jennings, 2012). Besides educational researchers, scholars from other fields such as psychology, economics, and sociology also provide their own classification or definition of SEC. Economists refer social and emotional behaviors as non-cognitive skills or personality traits (Borghans, Duckworth, Heckman, & Weel, 2008; Heckman & Rubenstein, 2001). Psychologists, on the other hand, define these skills as self-regulation indicators including attention shifting, effortful control, inhibitory control (Blair & Razza, 2007; Blair, 2002; Li-Grining, Votruba-Drzal, Maldonado-Carreno, & Haas, 2010), and examine their predictive relationship with internalizing and externalizing behaviors. Finally, sociologists describe these skills as learning-related behaviors and student engagement (Bodovski & Farkas, 2007; Mclelland et al, 2006; Morgan, Farkas, & Wu, 2001; Stipek et al, 2010). Despite the various terms used in the extant literature, the researchers across disciplines agree that there are associations between socioemotional competence and academic skills such as, reading, math, and science skills.

In summary, a fair amount of evidence has demonstrated that socioemotional competence is associated with reading achievement, and revealed the disparities relating to SES, gender, and ethnicity in both domains. However, few studies have examined the association between HLE and SEC, and the socioemotional development over time. I expect that there would be a SEC transition over time, and its long-term effect on the reading performance of students.

Review on Methodology

In this section, I shall first briefly introduce the three Latent Variable Modeling Methods separately in terms of the model building, model selection, and model interpretation as well as the Decision Tree. Then, I explain the rationale why I employ these two methods in my study.

Latent Variable Modeling is an umbrella term which subsumes latent class analysis, latent trait analysis (or item, response theory), latent profile analysis, and factor analysis. Further, it also has its longitudinal extension model – latent transition analysis (LTA). In this study, the latent class analysis, latent profile analysis and their longitudinal extension – latent transition modeling are used as the main analytic tools.

Latent class analysis. Latent Class Analysis (LCA) is a statistical tool for identifying underlying subtypes present in empirical data. “Latent” refers to the identified subtypes that are not measured directly, but indirectly by observed variables or manifest indicators (Collins & Lanza, 2010). Moreover, the identified subgroups bear similarity within groups, but differ between groups. In other words, LCA partitions the heterogeneous data patterns into subgroups with homogenous subgroups.

In the Latent Class Model, the latent variables and indicators are categorical, while when the indicator variables are only continuous, it is referred to as the Latent Profile Model and the corresponding analysis approaches are termed as Latent Profile Analysis (Vermunt & Magidson, 2002). In the first study, the LCA is the analytic tool, and the LPA is used in the second study.

A wide range of research questions can be answered by using this methodology. We can invoke LCA to discover the underlying classes, and further obtain the typical or salient qualitative features of subgroups, or compare the mean differences of certain measures or intervention conditions across latent classes. (Clark & Muthén, 2009) The estimated categorical latent variable can also serve as the moderator or mediator in the follow-up analysis. (e.g., Asparouhov & Muthén, 2014; Herman, Ostrander, Walkup, Silva, & March, 2007).

Parameters estimates. In the LCA, two types of parameters are usually estimated: the probabilities of membership in each latent class or class prevalence and the item response probabilities. In the LCA, cases (or observations, participants) are grouped into discrete latent classes. The model calculates the probabilities of membership in all estimated latent classes, and each individual fractionally belongs to all estimated latent classes simultaneously. As a result, the probabilities of all the estimated latent classes are summed to one since each class is mutually exclusive. It also reflects the uncertainty of classification for each individual. Because LCA can capture such uncertainty (analogous to the measurement error in SEM), the latent class results, in fact, were corrected for the uncertainty or error (Berlin, Williams, & Parra, 2013; Muthén, 2001; Nylund, Asparoushov & Muthén, 2007;).

Item response probability is the conditional probability of one single observed response pattern (e.g. endorsement of 1 or other certain value) on one particular indicator (item, measure, or task) conditional on one of the latent class classification results. It can be used to interpret the latent class solution. The item probabilities within each estimated latent class are also summed to one. (Collins & Lanza, 2010).

Estimation procedure. In this section, I introduce the basic steps to conduct LCA. The LPA has similar steps but with a few exceptions, which will be detailed in the LPA section. Available software for latent class analysis includes Latent GOLD (Vermunt & Magidson, 2005), Mplus (Muthén & Muthén, 1998-2012), and SAS PROC LCA/LTA (Lanza, Collins, Lemmon, & Schafer, 2007). In the current study, the Latent Gold and Mplus are used for the analysis. Latent Gold is used for the model exploration, and Mplus is used for the complex model building. The results of these two software analyses also serve the validation.

Data preparation and exploration. I examine the data and determine the appropriate analytic model. Several aspects have to be considered, namely: 1) the sampling procedure and data structure: whether the data has multiple levels; 2) the missing patterns: whether there exist any missing values (and how to handle the missing values; and 3) the scale of indicator: whether the indicator variables are categorical, or interval or mixture of two.

Model building. The most challenging step in the application of LCA is to determine the optimal number of the latent class, because the number is unknown. Hence, the model building is basically an exploratory procedure by enumerating the number of the latent class one by one, which is arduous and cumbersome. Ideally, the decision should be based upon both previous research and theory. However, as there is often minimal existing reading and literacy research using LCA and other Latent Variable Modeling, researchers can estimate multiple numbers of classes as long as the models can be statistically identified and practically interpreted (Berlin et al., 2013). The model building starts from 1-class to n-class till the model cannot be successfully converged and identified. In the meanwhile, researchers should document the mode fit information and other statistical results. Then researchers evaluate the candidate models based on the absolute model fit, relative model fit, classification solution and the substantive meaning. (Masyn, 2013)

Model selection and evaluation. Researchers can evaluate the candidate models and then identify the "optimal" fitting model based on the main categories: the absolute model fit, the relative model fit, the classification solution, and the substantive interpretation.

The absolute model fit refers to the likelihood ratio model chi-square goodness-of-fit (LRT) for the model with categorical indicators. The absolute test statistic evaluates whether the model well represents the actual data, or the consistency between the data and model. The non-significant test result is expected.

The relative model fit includes Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC), and Consistent Akaike's information Criteria (CAIC) fit statistics that are often used to compare models, where a lower value for each is preferable, indicating a better model. The relative model fit evaluates whether the model is better than another model in representing the actual data. (Lanza et al, 2007). It is noteworthy that even though one model might be better than another, both models might have poor fit to the data. Hence, the relative model fit and absolute model fit should be consulted simultaneously.

In addition, the Lo-Mendel-Rubin likelihood ratio test (LMR) and Bootstrap likelihood ratio test (BLRT) can be used to compare the improvement between neighboring models. For each of these tests, a *p* value less than .05 would indicate that the model with more classes (n+1 vs. n) would be a better fit to the data (Berlin et al., 2013). In brief, researchers should weigh model fit indices against the research questions and substantiate their findings when selecting the "best" model (Bauer & Curran, 2003).

As for the classification solution, the posterior item probabilities and the relative entropy statistic in R-squared (range: 0 to 1) can be used to evaluate the classification quality of cases in each model (Landa, Gross, Stuart, & Bauman, 2012). For example, for the 2-class model, when the posterior item probabilities on one item are far different from each other (0.9 vs. 0.1). The latent class well classifies the case on this item;

otherwise, for example, the posterior probabilities (0.5 v. 0.5), the latent class does not well separate the cases.

Model cross-validation. To verify and strengthen the validity of the class solution, I also conducted a double-cross-validation on the latent class solution in the first study based on the statistical results from both training and validation subsample. The steps are adopted from the study by Masyn (2013) with modification:

Step 1. Conduct the class enumerations from 1 to k till the model cannot be well identified.

Step 2. Determine the candidate models based on the model fit indices and classification quality. The candidate model should have adequate model overall fits (i.e., the LRTs statistical test results should not have statistically significant results).

Step 3. Conduct the cross-validation study on the candidate models. Randomly partition the entire sample into two subsamples with approximately equal sample sizes form Subsample A for data training and Subsample B for data validation (In the second round, the roles of each subsample are reversed: Subsample A for data validation, and Subsample B for data training).

Step 4. Fit candidate Latent Class Analysis Models to Subsample A (training dataset). Document the model fit indices and parameter estimates. Then use the estimate of the same candidate Latent Class Analysis Models to Subsample B. Instead of freely estimating the parameter meters, the parameters are fixed based on the estimates from Subsample A. If there is no evidence of the lack of the overall fit, then the model was validated well.

Step 5. Fit candidate models to Subsample B. At this time, the subsample B serves as the training dataset, and subsample B as the validation data set. Repeat the steps in Step 4. If the results do not indicate any lack of the overall fits, then the selection of this model is validated.

Step 6. If several candidate models validate well via the double cross-validation approach, then the more parsimonious model is preferred, that is, the model with the smaller latent class. It is noteworthy that latent class separation and substantive interpretation should also be considered.

Step 7. When the final model is decided based upon the cross-validation approach, this unconditional measurement model will be invoked for the following studies.

Latent class analysis with auxiliary (latent regression model) and distal outcomes. In this section, I briefly illustrate the estimation methods in the Latent Regression Model (Latent Class Analysis with predictors) and the Model with distal outcomes. In the present study, both extensions of the Latent Class Analysis are used for data analysis. Hence, it is necessary to give brief reviews from the conceptual and practical perspectives since they have not been widely used in reading and literacy research fields. However, the technical details about the model estimations are beyond the scope of this dissertation.

Estimation methods. Traditionally, researchers have used the classical three-step approaches to investigate the association between the predictors and the latent categorical variable. The classical approaches include step1: estimate the unconditional

LCA measurement model; step 2 obtain the estimated posterior probabilities, and assign the class membership based on the highest likelihood, while the variable of classification solution is also obtained. Step 3. The obtained classification variables would serve as outcome variables (e.g., in multicategorical logit model) or grouping variables (e.g., ANOVA or Chi-squared test) in the subsequent analysis (Feingold, Tiberio, & Capaldi, 2014, p 3). However, several simulation studies have the convergent findings that such approach underestimates the strength of association between the latent class variable and its observed predictor or outcome variables because such approach ignores the classification error, and the classification solution might be changed due to the covariates (Bolck, Croon, & Hagenaars, 2004; Vermunt, 2010 citied in Feingold et al 2014).

In the current study, I employed the 3-step approach and the BCH method proposed by (Vermunt, 2010; Asparouhov & Muthén, 2014 to correct for the classification error. The 3-step approach can be easily utilized in Mplus using the AUXILIARY option. For example, when SES (covariate) is added as a covariate, one can simply put the “R3STEP” in the parenthesis right after “SES” in order to specify the “SES” as the covariate (or latent class predictor) and the estimation method as the 3-step approach simultaneously and automatically. Such approach will not change the latent structure. The sample size in the analysis might shrink due to the missing values. The details have been discussed in Asparouhov and Muthén (2014)’ s study.

The modified BCH approach in practice for the model with distal outcome. In my study, I estimated the auxiliary regression model of the distal outcome on the covariates. To be more specific, I examined how the association between the dependent variable Y and several predictor variables Xs was moderated by the latent class. As a result, the latent class specific association can be estimated in the model. First, I estimated the unconditional LCA measurement model. During this step, the Y and Xs are specified in the AUXILARY Option to obtain the BCH weights dataset. Second, I estimated the Auxiliary model with distal outcomes based on the BCH weights dataset by incorporating the BCH weights obtained previously to correct for the errors (Asparouhov & Muthén, 2014).

Latent profile analysis.

***Parameter estimates*.** Different from LCA, the indicators in LPA are scaled continuously. Hence, the parameter estimates in LPA are the set of means, variances, and covariances within each latent class, in addition to the class prevalence rather than the item probabilities in LCA. Similar to the function of item probabilities, the means of items are usually used to interpret the latent class. Thus, the main goal of LPA is to identify the classes that differ regarding their means or locations (Vermunt & Magidson, 2002). When variances of each item are allowed to vary across classes, it implies that the classes may also differ concerning the homogeneity of the response to the observed variables.

***Restriction on parameter estimation*.** However, the number of free parameters will increase exponentially, as the number of indicators and latent class increase, especially in the unrestricted – class varying model. Hence, it is necessary to impose the constraints on the parameters for the successful model identification and stability and model parsimony and interpretability. In this section, the four main types of the variance-covariance matrix are discussed in the model estimation.

In LPA, the means of the set of the observed items are freely estimated across latent classes. The constraints are usually imposed on the variance and covariance matrix within and across latent classes. The restrictions usually can be applied to two dimensions: whether structure variability across latent classes: class-invariant or class-varying and whether covariances are freely estimated or not: unrestricted or diagonal.

The simplest variance-covariance matrix is the class-invariant-diagonal structure. In this model, only the means are freely estimated, but the variances are constrained to be invariant across latent classes. The diagonal variance-covariance matrix refers to the covariance among the observed items that were fixed to zero, namely, they are not estimated.

The next more slightly flexible model is the class-varying-diagonal structure. In this model, the means and variances are freely estimated across the latent classes. The latent class interpretations are based on both means and variances of the observed items. The covariances among the observed times are fixed to zero.

The third model is the class-invariant-unrestricted model. This model allows the variance-covariance matrix to be freely estimated, but the parameters have to be constrained to equivalent across latent classes.

The fourth model is the class-varying-unrestricted model. This model allows variance and covariance to be freely estimated and vary across latent classes. This model is the least restrictive. The model interpretation is based on the means, variances, and correlations among the observed items (Vermunt & Magidson 2002; Collins & Lanza; 2010).

Partial invariance model also can be tested, however, there is possibly infinite number of partially constrained models between the diagonal and the unrestricted model. The possible number will dramatically increase as the number of the latent class and indicator increases. Hence, in this study, I only examined the unconditional measurement model of LPA based on these four types of variance-covariance structure for exploratory purposes. LPA has the same steps as the LCA in the model building and model selection and evaluation.

Latent transition analysis. LTA is the longitudinal extension of the Latent Class/Profile Analysis, which models the transition of class membership over time. In LTA, the class membership might change in latent class over time. Namely, the classification membership is not assumed to be stable over time.

Parameter estimates. If the LTA model is based on the LCA model, three types of parameters are estimated. First, the latent class proportions (membership probabilities, or class prevalence) are estimated on each time points. Second, the transition probabilities over time points reflect the probabilities of moving from a particular latent class to another latent class at the next time points and the probabilities of staying in the similar latent classes over time (Lanza, Partick, & Maggs, 2010). Hence, Latent Transition Analysis also refers to Latent Mover-Stayer Model. Third, the item response probabilities reflect the conditional probabilities of endorsing one particular response pattern conditional on one particular latent class at each time points. In LTA, item response probabilities also are used to characterize the latent class. The first two types of parameters are structural parameters, and the item response probabilities are measurement parameters. If the LTA model is extended from the LPA model, instead of the item-response probabilities, three sets of measurement parameters are estimated: the means, variances, and variance-covariance matrix.

Restrictions on parameter estimates**.** In LTA, the restriction on parameters estimates mainly refers to the measurement invariance across time points**.** Restrictions can be imposed on both structural parameters and measurement parameters across the time points to obtain the measurement invariance over time. The measurement invariance over time has both conceptual and practical rationale. As for the conceptual rationale, for example in LCA, when item-response probabilities are constrained to be equivalent over time, this means the features of the latent class solutions are identical over time. The transition between the time points can be explained only in terms of the change in class membership or in the size of the latent class. While if the item-response probabilities are not identical across time points, the interpretation of the transition will be intricate due the different meaning of the latent class over time (Collins & Lanza, 2010).

Applied to LPA-based LTA, in addition to the restriction option on the set of means, they can be also applied to the variance and variance-covariance matrices. The measurement invariance can be obtained by constraining the means to be equivalent over time, and by restricting the variance and covariance matrices, or just variance, when in the diagonal model. The LPA-based LTA have much more flexibility than the LCA-based LTA, which also increases the uncertainty of the identification.

As for the practical rationale, the measurement invariance can facilitate the model identification and stabilize the model estimation. As Collins and Lanza (2010, p212) argue that "… it is a good idea to constrain the item-response probabilities in LTA to be equal across time whenever it is reasonable to do so, that is, whenever the measurement invariance across time can be reasonably assumed…". Likewise, it is also applied to LPA-based LTA. In the current study, the LTA is based on the LPA model. In terms of interpretation, the latent classes would be interpreted in terms of their means on each continuous indicator when modeling LTA on continuous scales. However, there are still no census steps to conduct the LTA on the continuous variables. More simulation study needs to be conducted in this field.

Decision tree. Decision Tree is a recursive partitioning procedure which was introduced by Breinman, Friedman, Stone, and Olshen (1984).The concept underlying this method is that it partitions the dataset into smaller and smaller homogeneous groups and fits model to the data as well as possible (Gordon, 2013). For the categorical response variable, the researchers intend to classify cases into groups by the categories or levels of the response and classification tree for this case. When the response variables are continuous, researchers intend to predict the variables, and the regression tree for this case. To be more specific, a classification tree splits the data based on homogeneity, classifies based on the similarity, minimizes the noise by "tree pruning". Instead, where the response variable is continuous, a regression model is fit to each of the predictors, isolating those variables as nodes where their inclusion minimizes the error (Morgan, 2014).

Growing decision tree. In this section, I introduce essential procedures in growing Decision Tree including splitting rules and model selection criterion. All the analysis was conducted in the SAS Enterprise Miner® (SAS Institute Inc., 2015).

First, I examined the statistical properties of each variable using the StatExplore node, then partitioned the input data into training and validation data set using Data Partition node. The Control Points node can simplify the data flow distribution in a more concise and clear format. The node Comparison is used for the model selection by comparing the fit indices based on the validation data results.

Splitting rules. The splitting rules determine which variables are used to define a split and what rule dictates how to conduct the split (see the SAS help manual). If the target variable is interval scaled, the SAS EM has ProbF and Variance options for splitting rules. For nominal targets, the splitting rules have three options: Gini, Entropy, and ProbChisq. The default setting for ProbChisq is 0.2. A split must be statistically significant at this level. Because the study used the binary response for the classification purpose, hence the splitting rules for this study are ProbChiq, Gini, and Entropy.

Model selection. Three models were computed based on the three splitting rules. The model selection is based on the validation misclassification rate. The model with smallest value indicates that it has the best fit to the data.

Why use person-centered analysis? The variable-centered methods such as regression analysis and factor analysis describe the associations between variables, while the person-centered analysis aims to identify the groups or clusters of individuals who share particular attributes within that same group. In addition, the assumptions of both methods are also different. The variable-centered approaches assume that the population is homogenous in regards to how predictors operate on outcome variables. Conversely, the person-centered methods (e.g., cluster analysis, latent class analysis, and latent transition analysis) assume that the population is heterogeneous. Finally, these two types of methods answer different sorts of research questions. The variable-centered approaches are used to answer how much independent variables can explain the variance in the outcome variables. The person-centered methods, by contrast, provide answers about what are the underlying types or profiles of certain groups and how they are associated with other variables (Laursen & Hoff, 2006; Magnusson, 2003, Muthén & Muthén, 2000).

In the current study, I attempt to explore the profiles of HLE and socioemotional competence among children at the early grade levels, and how the profiles or over-time transitions might differ by other variables and how much difference in certain variables may be attributed to the profiles. Hence, the person-centered analysis is well suited for the proposed study.

Why use decision tree? Decision Tree is a type of predictive modeling that presents results in a tree format which is easier for practitioners. Using Decision Tree, researchers can investigate large scale datasets, explicating previously underlying associations among variables. It is also a useful exploratory tool capable of developing a system to determine at-risk students (Koon, Petscher, & Foorman, 2014).

Different from the traditional methods (parametric approach such as linear regression), the Decision Tree model, a nonparametric and non-linear approach, does not require distributional assumptions. Neither does this method require any functional form for the predictors. Moreover, this approach has no assumption of additivity of predictors, thus allowing the researchers to identify the complex interactions among variables (Cordon, 2013, Kuhn, Page, Ward, & Worrall-Carter, 2014).

In contrast to the widely used traditional analysis methods, Decision is an emerging method in educational research, along with the growing numbers of researches applying this method in the Learning Analytics/Education Data Mining field. However, there is still a very limited use in the education research, let alone in the study of reading and literacy. For example, Compton, Fuchs, Fuchs, and Byrant (2006) employed Decision Tree to identify at-risk readers. They might have been the first who introduced this method into the field of reading research. Since then very few published studies used this method. In the third study, I attempted to identify the at-risk readers by using the two obtained latent discrete variables (i.e., HLE and SEC), the obtained transition response variables, and other family characteristics. The method allows me to model the complex interactions of those factors. Hence, this method is well suited to the purpose of my study.

CHAPTER III

METHODOLOGY

Data

The data for this study were extracted from the Early Childhood Longitudinal Study—Kindergarten Cohort (ECLS-K: 2011), which was sponsored by the U.S. Department of Education, National Center for Education Statistics. The ECLS-K 2011 collects information about the early educational experiences of a nationally representative sample. The data collection began in the 2010-2011 when the sampled children were in the kindergarten, and continued through the spring 2016, when children were promoted to the fifth grade. And this study only used the first four data points.

The currently released version is a nationally representative sample dataset including the information about the early educational experience from the kindergarten to the first grade. The dataset includes measures ranging from school and community factors to student-level factors regarding academic achievement, socio-economic status, social-skills and also information associated with their teachers and parents (Tourangeau et al., 2013).

Sample

The sample in this study includes students in the kindergarten for the first time in 2010-2011 school year and advanced to the Grade 1 in the next year. However, the time points for the current study are Kindergarten Fall and Grade 1 Spring semester. The full sample size of the ECLSK-2010-2011 is 18,174. After the data cleaning, the remaining sample size for the analysis is 13,367. The data preparation is detailed in the following section.

Measures

Reading achievement. The reading assessment includes questions measuring the basic skills such as print familiarity, letter recognition, beginning and ending sounds, recognition of the common words (sight vocabulary), and decoding multisyllabic words; vocabulary knowledge such as receptive vocabulary and vocabulary-in-context (Tourangeau et al., 2013). Since the item-level questions have not been released, the IRT-scaled reading scores will be used for the analysis, which means the emergent literacy skills cannot be examined for now.

Home literacy environment. The item-level questions about the home environments in the parent interview questionnaires can be used for the analysis. Specially, the items collected in the kindergarten fall 2010 will be utilized to identify the early Home Literacy Environment. The selected measures and codes are detailed in Table 2.

Socioemotional measures. Socioemotional development assessments includes multiple aspects of social skills (e.g., social interaction, attentional focus, and self-control), internalizing and externalizing problem behavior, and approaches to learning (i.e., keeps belongings organized; shows eagerness to learn new things; works independently; easily adapts to changes in routine; persists in completing tasks; pays attention well; and follows classroom rules). Both teachers and parents were asked to rate the children's social and emotional behavior. The analysis will be conducted based on the rating from both parents and teachers.

Approaches to learning items indicated a selected set of learning behaviors: keeps belongings organized; shows eagerness to learn new things, works independently, easily adapts to changes in routine, persists in completing task, pays attention well, and follows classroom rules. Higher scale scores indicate that the child exhibited positive learning behavior more often (Tourangeau et al., 2013). The selected measures and codes are detailed in Table 3.

Table 2

Home Literacy Environment Measures in ECLS-K: 2011

Description	Question Code	Cohort	Variable Code
How often parents tell stories to child	HEQ010A	P1	TELLST
How often parents sing songs with child	HEQ010B	P1	SINGSO
How often parents help child do art	HEQ010C	P1	HLPART
The frequency of involving child in household chores	HEQ010D	P1	CHORES
How often parents play games or do puzzle with kids	HEQ010E	P1	GAMES
The frequency of talking about nature	HEQ010F	P1	NATURE
You all build things	HEQ010G	P1	BUILD
You all do sports	HEQ010H	P1	SPORT
Practice reading, writing or working with number	HEQ010I	P1	NUMBERS
Read books to child	HEQ030	P1	READBK
How long read book to child? (mins)	HEQ036	P1	RDMINS
How many books child has	HEQ040	P1	CHILDBK
How often child reads picture books	HEQ060	P1	PICBK
How often child reads outside of schools	HEQ070	P1	CHREAD

Table 3

Socioemotional Competence Measures in ECLS-K: 2011

Description	Cohort	Variable Code
Parent report Impulsive/Overactive	X1	PRDNIMP
Parent report approaches to learning	X1	PRNAPP
Parent report self-control	X1	PRNCON
Parent report sad/lonely	X1	PRNSAD
Parent report social interaction	X1	PRNSOC
Teacher report approaches to learning	X1	TCHAPP
Teacher report self-control	X1	TCHCON
Teacher report externalizing problem behaviors	X1	TCHEXT
Teacher report internalizing problem behaviors	X1	TCHINT
Teacher report interpersonal skills	X1	TCHPER
Parent report Impulsive/Overactive	X4	PRDNIMP
Parent report self-control	X4	PRNCON
Parent report sad/lonely	X4	PRNSAD
Parent report social interaction	X4	PRNSOC
Parent report approaches to learning	X4	PRNAPP
Teacher report self-control	X4	TCHCON
Teacher report externalizing problem behaviors	X4	TCHEXT
Teacher report internalizing problem behaviors	X4	TCHINT
Teacher report interpersonal skills	X4	TCHPER
Teacher report approaches to learning	X4	TCHAPP

SES. SES measures include the components of father/male guardian's education, mother/female guardian's education, father/male guardian's occupational prestige, mother/female guardian's occupational prestige, and household income. (Tourangeau et al., 2013). The composite variable will be used for the analysis (X12SESL).

Race/Ethnicity. The race/ethnicity variables are created based on the composite variable X_RACETHP_R as the original dichotomous variable cannot be directly used in the present study (X_HISP_R, X_AMINAN_R, X_ASIAN_R, X_HAWPI_R, X_AFRICAN AMERICAN_R, X_WHITE_R), because those variables cannot uniquely identify the ethnicity group. The new dichotomous variables were labelled as D1, D2, D3 in the analysis (see Table 4)

Table 4

Dummy Coding for Race/Ethnicity

	D1	D2	D3
Caucasian	0	0	0
African American	1	0	0
Hispanic	0	1	0
Asian	0	0	1

Gender/Sex. Information about child's gender was collected from parents in the fall parent interview, and confirmed by parents in the spring parent interview (X_CHSEX).

Analytical Strategy

The remaining sections illustrate the analysis plans for each study and also the data preparation procedure.

What are the types of HLE profile of students at the kindergarten entry (Kindergarten, Fall 2010)? The scope of the HLE profile dimension is determined by the availability of the items in the ECLS-K dataset and extant literatures. The final resulting variables in this study include 13 items. Preliminary latent class was conducted on the original scale (1-4). The scale was treated as nominal and ordinal. However, neither situation allows for the successful cross-validation for the unconditional measurement model. Hence, the original 4 points scale is dichotomized from a new list of indicator for latent class analysis. The original scale for Home Environment, Activities and Cognitive Stimulation (HEQ) is from 1 to 4, with the coding scheme of 1 = not at all, 2 = once or twice a week, 3 = three to six times a week, and 4 = every day. The recoded variable collapsed 1 and 2 in the original scale into 1, and 3 and 4 in the original scale into 2. The recoding step is necessary for the successful model identification and model cross-validation and easier interpretation of the latent class results. Moreover, the PRIDMINS was removed from the analysis, because the range of the P1RDMINS is quite wide, which might be due to the misconceptions of the responders Moreover, the preliminary analysis also indicated the P1RDMINS cannot well result in a clear class separation in either the points scale or dichotomous scale. The new variables were labelled with the initial letters of N2. The thirteen items measure different aspects of home literacy environment directly and indirectly. The indirect measures include those items that reflect the literacy interaction via the parent-children collaborations or interaction such as doing sports, singing songs, playing games, doing households chores together. The literacy interaction occurred during the informal settings. The direct home literacy environment measure reflects the direct and explicit literacy. For example, telling the stories, reading books to kids, reading picture books, etc. Hence, the entire 13 items comprehensively, even though not exhaustively, measure the multi-dimensional concepts of the home literacy environment.

Step 1 model building. During the model enumeration stage, the latent class modes are estimated from class 1 to class 6 on the 13 binary home literacy environment items on the full sample data set to identify candidate fit models based on the non-statistically non-significant LRT results, and other relative fit indices.

Step 2 model cross-validation. To validate the selected candidate models, a double cross-validation analysis will be conducted on the selected unconditional measurement models. The dataset is randomly portioned into the training and validation data set. Then Latent Class Model is conducted on the training data, the obtained parameter estimates are used for the model estimation on the validation data set. The second run the model estimates on the validation data first, and estimate the model on the training dataset with the obtained parameters from the validation data. If both runs give the non-significant LRT result, the Latent Class Model with this particular latent structure is validated and supported. When multiple latent structures are supported, the more parsimonious one (fewer free parameters estimates, fewer latent class) or simplest model is preferred. The practical meaning is considered as well.

The final model is used to profile the home literacy environment at students' early school early. In this study, the time point is student's Kindergarten in Fall 2010.

Does the membership of identified HLE profile vary upon SES (i.e., family income, maternal education) and membership in racial groups? The Latent Regression Analysis Model (i.e., the latent class analysis with predictor variable) based on the unconditional model obtained from the previous analysis was conducted in the present study. In this analysis, it was assumed that the predictors or covariates do not shift the latent class solution. In other words, the resulting latent class prevalence will not change depending on the impact of the covariates. Instead, how the latent class prevalence varies upon the participant's SES and their ethnicity group was inspected.

The analysis in this study is basically analogous to the mutli-nominal logistic regression. Differently, the dependent variable in this analytic model is latent or unobserved. To correct for classification and measurement errors, I employed the new-3-Step approach (Asparouhov & Muthén, 2014) to estimate the Latent Regression Mode by adding R3STEP in the Auxiliary Variable option in Mplus which results in an automatic new 3-step approach estimation (as noted in the previous section). First, the model with the latent class predictors: ethnicity and SES, respectively was estimated. Next, the model with both latent class predictors together to identify the distribution of the latent class membership across ethnicity and SES was computed. The final results were visualized via graphic plots.

What is the association between HLE and students' early reading achievement after controlling for SES and racial groups? The purpose of this analysis was to examine the association between participants' demographic factors (SES and ethnicity) and their reading achievement. The dependent variables in this analysis are the IRT reading scores on fall kindergarten and spring grade 1. I conducted the Latent Class Analysis with both predictors and distal outcomes (i.e., Latent Regression Model with distal outcome [LRM-D]) to examine how such association is decomposed and varies across the home literacy profile. Three LRM-D models respectively were computed. The first LRM- D is the model with the fall kindergarten IRT score as the distal outcome. The second one is the model with the spring grade 1 IRT as the distal outcome. The third one is the model where the fall kindergarten IRT reading scores together with the two demographic factors as the covariate, and the spring grade 1 IRT reading score as the distal outcome.

What is the latent profile of SEC, and its variation in membership on covariates including SES and gender? The Latent Profiles of SEC were conducted for teacher and parent response for these two time points without adding covariates. The scale of the social and emotional competence was assumed to be continuous. Instead, the Latent Profile Analysis on the items of the SEC on both time points was computed. Multiple unconditional measurement models were estimated from 1-class to -6-class, the various variance-and covariance matrix structures also applied to the model estimation.

After obtaining the unconditional measurement model, I applied the Latent Regression Model to the data on both time points by adding covariate to the unconditional measurement model in order to examine the variation of SEC profile prevalence upon the covariates (SES and gender).

Whether the SEC profile is stable over time? And what are the long-term effects of SEC on later reading comprehension? To test the stability of the latent structures over these two times, I applied the Latent Transition Model to the sampled data. Conventionally, the LTA model would be based on the selected optimal unconditional LCA measurement models, and its basic estimated procedure (Nylund, 2007) for the LTA estimation is as follows:

Step 1. Investigate measurement model alternative for each points.

Step 2. Test for measurement invariance across times.

Step 3. Explore specification of the latent model without covariates.

Step 4. Include covariates in the LTA model.

Step 5. Include distal outcomes.

However, unlike the LTA model demonstrated by Masyn (2013), in the current study, the LTA model was built based on the LPA. As noted in the previous chapter, because the LPA has more flexible variance-covariance structure than LCA, it would result in much more candidate models. Hence, the analytic steps would be adjusted for it. The analytic steps for LTA in this analysis were as follows: .

Step 1. Estimate the LTA based on the unconditional measurement LPA. The obtained LTA model serves as the baseline model. The assumption was the measurement invariance over times for easier interpretation of latent structure, and transition patterns.

Step 2. Estimate other candidate LTA models with different variance-covariance matrix structures, while fixing the measurement invariance over time.

Step 3. Select the final optimal models based on the model fit indices, class separation, and substantive meaning.

Step 4. Include covariates in the LTA model and included SES, gender as the covariates to examine the latent transitional patterns vary upon on the different level of SES and gender.

Step 5. Include distal outcomes and added the fall 2011 IRT reading score as the distal outcome.

What is the variation in later reading achievement predicted by the profiles of the HLE and SEC, and SEC transition response patterns, as well as their interaction with other covariates? The Decision Tree on nominal target variable for this analysis was computed. First, the data set was partitioned into training and validation dataset. The analysis on the training data served to build the model, the analysis on the validation data served to select the model.

To run the classification tree, the IRT reading scores were dichotomized into binary variables based on the cut-off value (the bottom 10th percentile). The binary variable is the indicator of the at-risk readers and the results from this analysis served as the warning system to identify at-risk readers.

CHAPTER IV

RESULTS

This chapter presents the results of the study. The statistical modeling results that were conducted on the extracted data are reported in the four sections. The first section presents the results for study1 including the missing values information, descriptive statistics, Latent Class Analysis, cross-validation, Latent Regression Analysis, and Latent Regression with the distal outcomes. Those results were used to answer each research question in Study 1. Similar to the reported results in the first section, the second section, in addition, also reports the results of the Latent Transition Analysis, and LTA with the covariate and distal outcomes. The results in this section were used to answer each research question in the Study 2. The third section presents the results of the Decision Tree analysis. The results were used to build the at-risk readers identification system.

Study 1 Results

Missing values and missing patterns. The coding scheme for the Home Literacy Environment is shown in Table 5. The original data sample size was 18,174. The data extraction criterion is based on the missingness of all the home literacy environment HLE items. The data had 13,089 observations with complete information, while 4,807 observations had complete missingness. Table 5 also presents the codes of the missing values. The remaining 278 observations had the partial missingness on the HLE measures. To maximize the sample size in this study, 278 observations were included in the analysis together with the 13,089 observations. The final sample size for this study was 13,367. All the analyses were conducted in Mplus. Mplus has the FIML estimation option that would well handle the missing values in the dataset. Since FIML required the missing patterns to be either MAR or MCAR, Little's MCAR test was also conducted during the analysis.

Table 5

Coding Schemes for HLE Recoded Binary Indicators and Missing Values

Code	Response
-9	Not Ascertained
-8	Don' t know
-7	Refused
-1	Not Applicable
4	Every day
3	3 to 6 times a week
2	Once or Twice a week
1	Not al all
New coding	
2*	More than 3 times a week
1*	Less than 2 times a week

Descriptive analysis. The scales of the home literacy environment measures were rescaled into binary for the purposes of the model identification and interpretation. The rescaling details are displayed in the method section. Table 6 presents the univariate proportions and count for binary variable. Most parents are more likely to get involved with home literacy activities directly or indirectly, with some exceptions such as the item N2NATURE (i.e., How often you talk about nature) where most parents are unlikely to talk about it with children.

Table 6

Univariate Proportions and Counts for Binary Response of 13 Indicators

	Category	Proportions (%)	Counts	Missing observation	Missing proportions (%)
N2TELLST	1	29.3	3914	7	0.05
	2	70.7	9446		
N2SINGSO	1	27.6	3684	7	0.05
	2	72.4	9676		
N2HLPART	1	40.6	5416	11	0.08
	2	59.4	7940		
N2CHORES	1	22.3	2974	8	0.06
	2	77.7	10385		
N2GAMES	1	34.5	4615	9	0.07
	2	65.5	8743		
N2NATURE	1	65.8	8779	23	0.17
	2	34.2	4565		
N2BUILD	1	56.2	7501	20	0.15
	2	43.8	5846		
N2SPORT	1	38.4	5133	13	0.10
	2	61.6	8221		
N2NUMBRS	1	7.70	1029	13	0.10
	2	92.3	12325		
N2READBK	1	14.5	1940	15	0.11
	2	85.5	11412		
N2PICKBS	1	18.9	2515	37	0.28
	2	81.1	10815		
N2CHREAD	1	28.6	3806	38	0.28
	2	71.4	9523		
N2CHLDBK	1	53.8	7191	--	--
	2	46.2	6176		

Moreover, some other HLE measures have almost equal likelihood of response on both categories such as N2BUILD (How often you build things), N2CHILDBK (how many books a child has) and N2HLPART (how often you help children to do art).

Results of latent class analysis. Table 7 provides goodness-of-fit statistics for the one-class, two-class, three-class, four-class, and 5-class models on the full sample size data (n=13,367) of 13 binary variables. The reason why the class enumeration ends till 5-class is that the latent class solution does not display clear patterns. In addition, the Little's MCAR test was also reported. As shown, the LRT-CHSQ in 3-Class, 4-Class, and 5-Class models do not indicate statistically significant results at the alpha = 0.05 level. Moreover, the assumption of MCAR on the missingness also holds, indicating that the FIML methods are appropriate for this analysis.

Model selection. As shown in Table 7, the 3-Class, 4-Class and 5 Class indicated the adequate model fit, with $p > 0.99$ for all the three models. Moreover, the VLMLR test and LMR test showed the significant improvement in model fit from K-1 Class model to K-Class model. However, based upon the model parsimony, the 3-Class is preferred.

Model cross-validation. Table 8 presents the cross-validation results. The follow-up cross-validation analysis was conducted on the 3-Class unconditional measurement model. The entire sample was randomly partitioned into two subsamples: A and B. During the training step, Sub-Sample A had sample size of 6683, indicating the adequate model fit of G^2= 5366.808, df = 8112, p > .99. So do the results of the Sub-Sample B (G^2=5449.108, df = 8112, p > .99). The validation step also indicated the model fit on both sub-samples when estimating the model with fixed parameters from the training steps. In summary, the results of the cross-validation corroborate the validation of the 3-Class Latent Class Model.

Table 7

Goodness of Fit, Model Comparison, and Little's MCAR Test

	1-class	2-class	3-class	4-class	5-class
n-par	13	27	41	55	69
-LL	-100090.306	-92962.675	-91531.965	-91171.107	-90968.061
AIC	200206.613	185737.349	183145.910	182452.214	182074.122
BIC	200304.120	186181.864	183453.433	182864.744	182591.660
SBIC	200262.807	186096.061	183323.139	182689.959	182372.384
LRT-CHSQ	12238.781	8491.155	7479.030	6720.530	6458.972
df	8049	8108	8100	8083	8070
p-value	<.05	<.05	>.99	>.99	>.99
Entropy	NA	0.709	0.694	0.651	0.610
n-1 vs. n					
VLMLR test	NA	<.0001	<.0001	<.0001	0.0046
LMR-LRT test	NA	<.0001	<.0001	<.0001	0.0048
Little's MCAR test					
LRT-CHSQ	280.844	253.679	301.261	294.075	301.472
df	53149	53149	53149	53149	53149
p-value	>.99	>.99	>.99	>.99	>.99

Note: LL= likelihood, AIC = Akaike Information Criterion, BIC = Bayesian Information Criterion, SBIC= ScharzBIC, LRT-CHSQ = Likelihood Ratio Chi-Square test, VLMLR = Vuong-Lo-Mendell-Rubin test, LMR = Lo-Mendell-Rubin, MCAR = missing completely at random.

Table 8

Cross-Validation for 3-Class Latent Class Analysis

SubSampA	n	Parameters Restriction	n-par	LR-CHSQ	df	p
	6683	Free	41	5366.808	8112	>.99
		Fixed*	0	5463.645	8153	>.99
SubSampB						
	6684	Free	41	5449.108	8112	>.99
		Fixed+	0	5510.393	8153	>.99

Note: * fixed with subsample B parameters. + fixed with subsample A parameters.

What are the HLE profiles of students at kindergarten entry (Kindergarten, 2010 Fall)? Table 9 presents the item response probabilities and latent class prevalence of the 3-Class model for each latent class. Figure 6 visualizes the results for easier interpretation. As shown in the web graph, the Class 1 is characterized by the highest probabilities of endorsing the value of 2 for each item. As shown in the coding table, the value of 2 refers to more than 3 times for the particular home literacy behavior. Hence, the Class1 can be interpreted as the pro-reading group (named pro-reading1). Likewise, the Class2 also had the similar response patterns. The salient difference between Class1 and Class2 merely lies in the response of the item "NATURE" (the frequency of talking about nature with kids, the Class1:0.67 vs. Class 2: 0.20). Therefore, the Class2 was also rendered as the pro-reading group (named pro-reading 2). The Class 3 displayed the relatively low likelihood of the pro-reading response (i.e., response = 2; named contra-reading). As for the class prevalence, the Class 2 (pro-reading 2) is the predominant subgroup with the size of 6871 participants (51.4%), in comparison with the Class 1 (pro-reading1, n= 4536, 33.9%) and the Class 3 (contra-reading, n = 1960, 14.6%). Overall, approximately 85% of parents actively participated in the direct and non-direct literacy activities with their children based on the parents' self-reported home literacy environment survey items.

In conclusion, based on the survey items in Fall 2010, three latent classes to profile the home literacy environment were identified: the Class 1 (pro-reading1), the Class 2 (pro-reading 2), and the Class 3 (contra-reading). Most parents reported that they actively support their kids' reading by providing rich literacy resources (books), frequent interaction (i.e., reading book to kids, play games and sing songs with kids).

Table 9

Results of Latent Class Analysis: Latent Class Prevalence, Item Response Probabilities (Category =2)

Class	Class 1	Class 2	Class 3
n	4536	6871	1960
proportion	0.339	0.514	0.146
TELLST	0.94	0.71	0.20
SINGSO	0.93	0.68	0.43
HLPART	0.90	0.49	0.29
CHORES	0.94	0.75	0.51
GAMES	0.95	0.57	0.32
NATURE	0.67	0.20	0.13
BUILD	0.72	0.33	0.19
SPORT	0.87	0.54	0.34
NUMBRS	0.99	0.95	0.70
READBK	0.99	0.93	0.34
PICBKS	0.95	0.86	0.35
CHREAD	0.91	0.72	0.28
CHLDBK	0.62	0.46	0.14

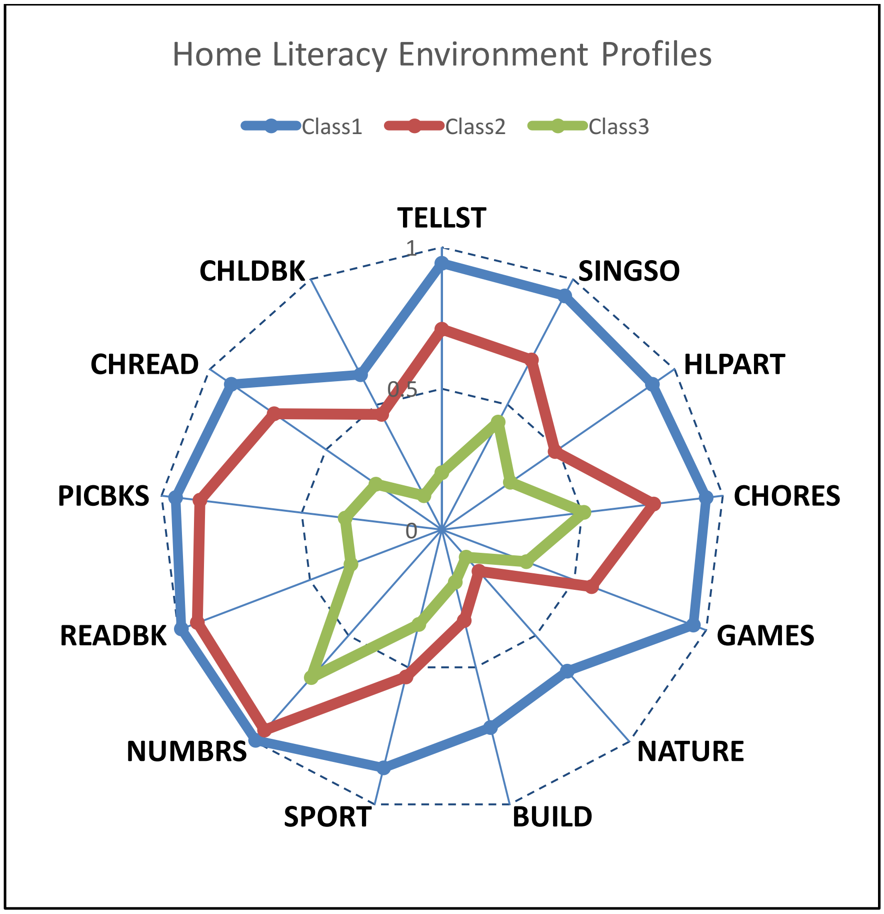


Figure 6 Plot of Item Response Probabilities for Identified Latent Class

Does the membership of the identified HLE profiles vary upon racial groups and SES (i.e., family income, maternal education) and membership in racial groups? First, Latent Regression Analysis with the covariate of race was conducted to explicate the association between the HLE profiles and race. Table 10 displays the logit estimates. These parameters are estimates from the multi-nominal logit regression with the correction for classification error within the latent regress model. The racial group variables are three dummy coded indicators with Caucasian as the reference group and the analytic model invoked Caucasian group as the baseline. Compared with the Caucasian group, all the other three racial groups showed significant difference in terms of the relative membership of the certain latent class. For example, the odds ratio value of 0.237 means that the odd of African-Americans being in Class 1 rather than Class3 is 0.237 times that for Caucasian. In another words, Caucasians families had 4.21times (1/0.237) likelihood of belonging to Class 1 over Class 3 than African Americans.

For easier interpretation, the estimated logit parameters were converted into probabilities. Table 11 presents the latent class prevalence across racial groups. Across four racial groups, the likelihood of being classified as Class 2 is approximately .5. Similar to the unconditional LCA, the Class 2 is the prevalent latent class, and the combined likelihood of Class 1 and Class 2 are above .7 across racial groups. Caucasian group had the probability of .926 to be classified as pro-reading group. The likelihood for African-American group was .788; for Asians the value was .734, and Hispanic parents had the lower probabilities to participate in children's literacy activities (.704).

Table 10

Parameter Estimates of Latent Class Categories by Racial Groups

(CLS1/CLS3)	Estimates	s.e.	MOE	LL	UP	OR	LL	UP
Intercept	1.678	0.061	0.120	1.558	1.798			
D1 (African American/Caucasian)	-1.438	0.102	0.200	-1.638	-1.238	0.237	0.194	0.290
D2 (Hispanic/Caucasian)	-1.901	0.085	0.167	-2.068	-1.734	0.149	0.126	0.177
D3 (Asian/Caucasian)	-1.960	0.134	0.263	-2.223	-1.697	0.141	0.108	0.183
(CLS2/CLS3)								
Intercept	1.962	0.064	0.125	1.837	2.087			
D1 (African American/Caucasian)	-1.065	0.103	0.202	-1.267	-0.863	0.345	0.282	0.422
D2 (Hispanic/Caucasian)	-1.507	0.085	0.167	-1.674	-1.340	0.222	0.188	0.262
D3 (Asian/Caucasian)	-1.265	0.122	0.239	-1.504	-1.026	0.282	0.222	0.358
(CLS1/CLS2)								
Intercept	-0.284	0.034	0.067	-0.351	-0.217			
D1 (African American/Caucasian)	-0.372	0.085	0.167	-0.539	-0.205	0.689	0.584	0.814
D2 (Hispanic/Caucasian)	-0.393	0.071	0.139	-0.532	-0.254	0.675	0.587	0.776
D3 (Asian/Caucasian)	-0.695	0.127	0.249	-0.944	-0.446	0.499	0.389	0.640

Table 11

Estimated Probabilities of Latent Class Categories by Racial Groups

Race	Class 1	Class 2	Class 3
Caucasian	0.398	0.528	0.074
African American	0.269	0.519	0.212
Hispanic	0.237	0.467	0.296
Asian	0.200	0.534	0.266

In conclusion, regardless of the racial groups, most families provided supportive literacy environment for children. However, the salient racial group specific variation also exits, where Caucasians had the highest likelihood of supporting literacy activities, and other racial groups tended to have relatively low probabilities.

Second, to examine the association between the HLE profiles and racial groups controlling for SES, the variable SES was added as the second covariate to the previous latent regression model. Table 12 presents the multi-nominal logit parameters along with the odds ratio estimates. The variable SES showed significant effect in profiling HLE in terms of the relative latent class membership. For instance, the odds ratio value, 2.866, means when one SD increases in SES, the odds of being in Class 1 over Class 3 is increased by 2.866. Namely, the high SES family tends to more actively participate in the literacy interaction with children, regardless of the racial groups.

The estimated probabilities for each racial group at the particular SES levels are reported in Table 13 to 16 and Figure 7 to 10. Five SES levels (i.e., -2, -1, 0, 1, 2) were selected for the reported results. The Figures also display the combined likelihood of the two pro-reading groups. Across four racial groups, as SES increased, the likelihood of being in the pro-reading group grew significantly. In contrast, the likelihood of being in the contra-reading group plummeted sharply.

However, the critical turning points of the class membership likelihood located on different SES levels between racial groups. When the SES value was above the corresponding SES value, the likelihood of being in the pro-reading group gradually increased compared to that of the contra-reading group. When the SES value was below that value, the likelihood of belonging to the pro-reading group constantly decreased compared to that of the contra-reading group. The corresponding SES value for the Caucasians was approximately -2.5. For Hispanics, the value was -1.5. African-Americans had the value of -1.8. The Asian group had the corresponding SES value of -0.7. Hence, for the Asian family, the SES level seems to play a larger role in the parents' participation in literacy activities, in contrast to the situation in the Caucasian, Hispanic and African American families.

Table 12

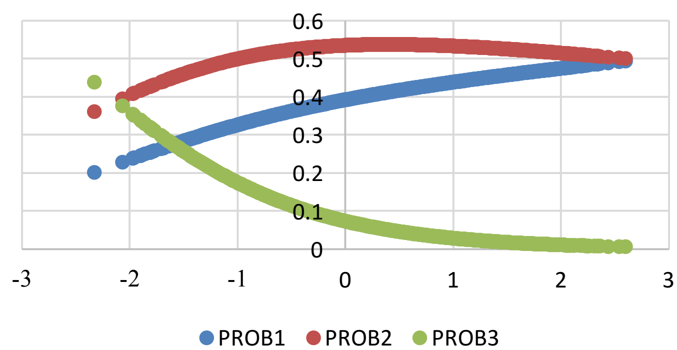
Parameter Estimates of Latent Class Categories by Racial Groups and SES

(CLS1/CLS3)	Estimates	s.e	MOE	LL	UP	OR	LL	UP
Intercept	1.679	0.064	0.125	1.554	1.804			
D1 (African American/Caucasian)	-0.948	0.060	0.118	-1.066	-0.830	0.388	0.345	0.436
D2 (Hispanic/Caucasian)	-1.296	0.089	0.174	-1.470	-1.122	0.274	0.230	0.326
D3 (Asian/Caucasian)	-2.323	0.143	0.280	-2.603	-2.043	0.098	0.074	0.130
SES	1.053	0.057	0.112	0.941	1.165	2.866	2.563	3.205
(CLS2/CLS3)								
Intercept	1.990	0.067	0.131	1.859	2.121	7.316	6.415	8.342
D1 (African American/Caucasian)	-0.628	0.110	0.216	-0.844	-0.412	0.534	0.430	0.662
D2 (Hispanic/Caucasian)	-0.978	0.090	0.176	-1.154	-0.802	0.376	0.315	0.449
D3 (Asian/Caucasian)	-1.576	0.131	0.257	-1.833	-1.319	0.207	0.160	0.267
SES	0.938	0.058	0.114	0.824	1.052	2.555	2.280	2.862
(CLS1/CLS2)								
Intercept	-0.311	0.035	0.069	-0.380	-0.242	0.733	0.684	0.785
D1 (African American/Caucasian)	-0.320	0.086	0.169	-0.489	-0.151	0.726	0.614	0.859
D2 (Hispanic/Caucasian)	-0.318	0.075	0.147	-0.465	-0.171	0.728	0.628	0.843
D3 (Asian/Caucasian)	-0.747	0.128	0.251	-0.998	-0.496	0.474	0.369	0.609
SES	0.114	0.037	0.073	0.041	0.187	1.121	1.042	1.205

Table 13

Estimated Probabilities for Caucasian Group by SES

SES	prob1	prob2	prob3	prob12	prob3
2	0.475	0.515	0.011	0.989	0.011
1	0.438	0.533	0.029	0.971	0.029
0	0.392	0.535	0.073	0.927	0.073
-1	0.326	0.499	0.174	0.826	0.174
-2	0.235	0.404	0.361	0.639	0.361



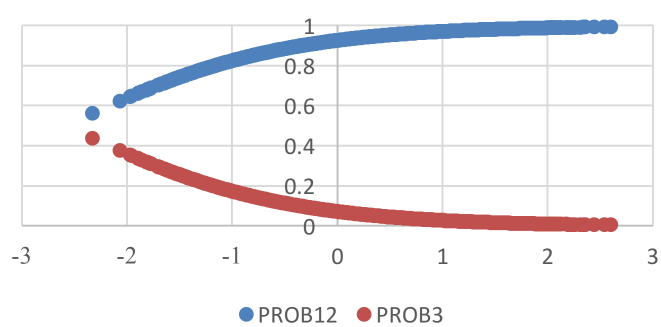
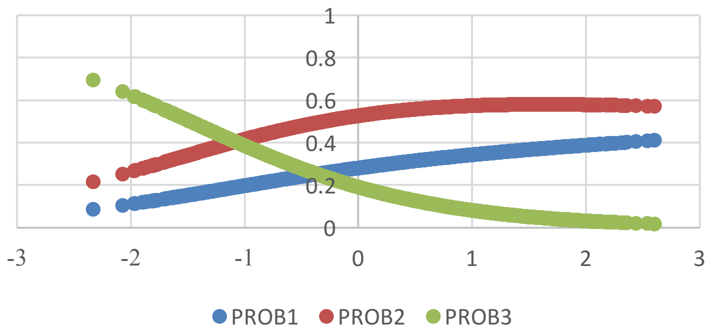


Figure 7 Plots for Estimated Probabilities and Combined Probabilities for Caucasian Group by SES.

Table 14

Estimated Probabilities for Hispanic Group by SES

SES	prob1	prob2	prob3	prob12	prob3
2	0.389	0.579	0.032	0.968	0.032
1	0.344	0.575	0.082	0.918	0.082
0	0.281	0.527	0.192	0.808	0.192
-1	0.198	0.416	0.386	0.614	0.386
-2	0.112	0.263	0.625	0.375	0.625



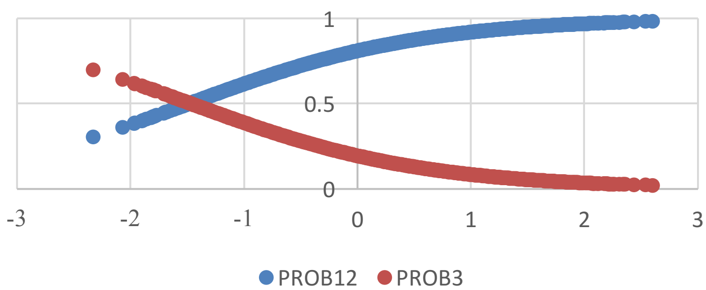
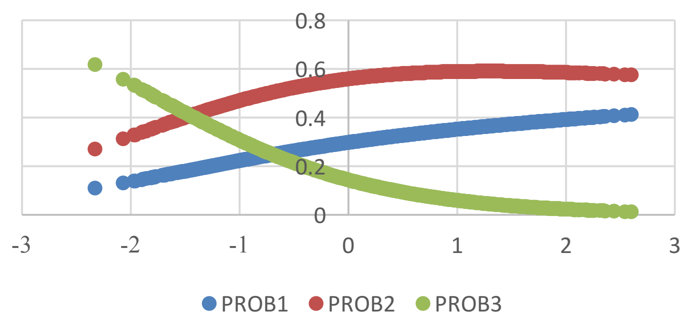


Figure 8 Plots for Estimated Probabilities and Combined Probabilities for Hispanic Group by SES

Table 15

Estimated Probabilities for African American Group by SES

SES	prob1	prob2	prob3	prob12	prob3
2	0.392	0.585	0.023	0.977	0.023
1	0.352	0.589	0.059	0.941	0.059
0	0.298	0.559	0.143	0.857	0.143
-1	0.223	0.470	0.307	0.693	0.307
-2	0.137	0.323	0.540	0.460	0.540



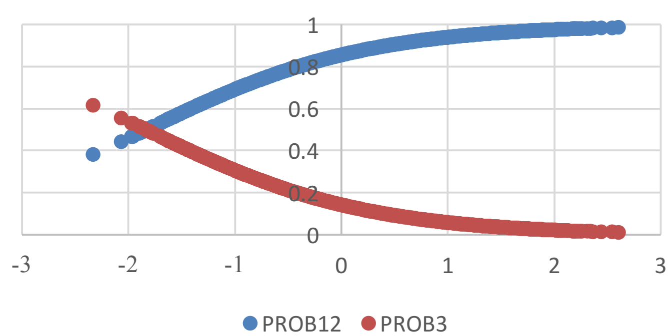
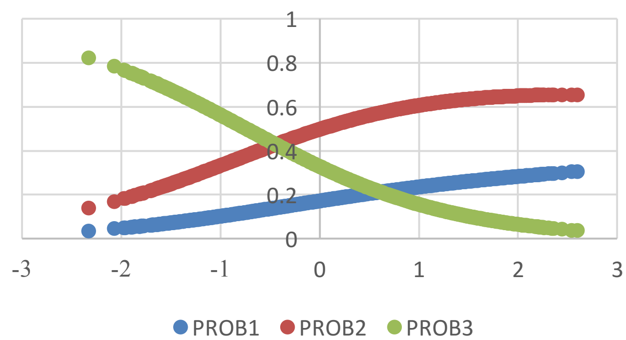


Figure 9 Plots for Estimated Probabilities and Combined Probabilities for African American Group by SES

Table 16

Estimated Probabilities for Asian Group by SES

SES	prob1	prob2	prob3	prob12	prob3
2	0.284	0.650	0.066	0.934	0.066
1	0.236	0.607	0.157	0.843	0.157
0	0.173	0.498	0.329	0.671	0.329
-1	0.103	0.334	0.563	0.437	0.563
-2	0.049	0.179	0.772	0.228	0.772



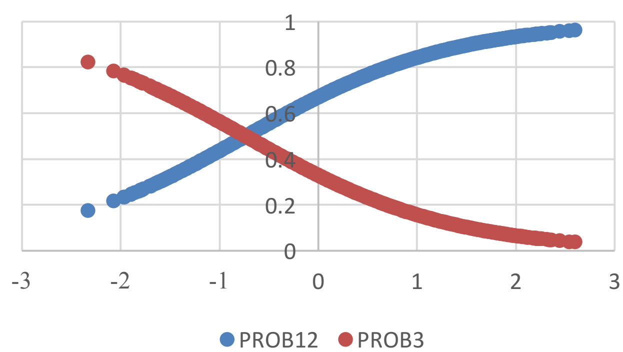


Figure 10 Plots for Estimated Probabilities and Combined Probabilities for Asian Group by SES

How does the association between the reading achievement and demographic factors vary upon the HLE profiles? Early reading achievements were measured by IRT reading achievement Fall Kindergarten 2010 and Spring grade1 2012. The Latent Regression Analysis with the distal outcome was conducted for both reading measures separately.

Results of the analytic models at fall kindergarten 2010. I conducted latent class analysis with the outcome on both reading measures respectively to examine the variation in the reading achievement across the HLE profiles. The estimated means for each latent class are displayed in Tables 17 and 18. At Fall Kindergarten 2010, the average reading achievement for each latent class are 38.787 for Class 1, 38.277 for Class 2, 33.914 for Class 3. There is no statistically significant difference between Class 1 and Class 2 (Chi-squared = 3.801, df =1, p = 0.051). However, the average reading achievement for Class 3 is far below Class1 (Chi-squared = 300.866, df =1, p <.005) and Class 2 (Chi-squared = 224.287, df = 1, p <. 005). Similar results were also found in the reading achievement at Spring Grade1 2012. The two pro-reading groups do not differ significantly from each other (Chi-square = 0.116, p = 0.733), but far above the Class 3 (Class1 vs. Class 3: Chi-square = 410.493, df =1, p<.005; Class2 vs. Class3: Chi-square = 363.724, df =1, p < .005). The visual comparisons of the corresponding confidence intervals are presented in Figures 11 and 12.

Table 17

Mean Estimates and Equality Tests of Means across Classes Using the BCH Procedure (Dependent Variable: X1RSCALK1-Fall 2010)

	Estimates	s.e.	MOE	LL	UL
CL1	38.787	0.179	0.351	38.436	39.138
CL2	38.277	0.149	0.292	37.985	38.569
CL3	33.914	0.223	0.437	33.477	34.351
	CHSQ	p-value			
Overall test	336.565	0.000			
CL1 vs. 2	3.801	0.051			
CL1 vs. 3	300.866	0.000			
CL2 vs. 3	224.287				

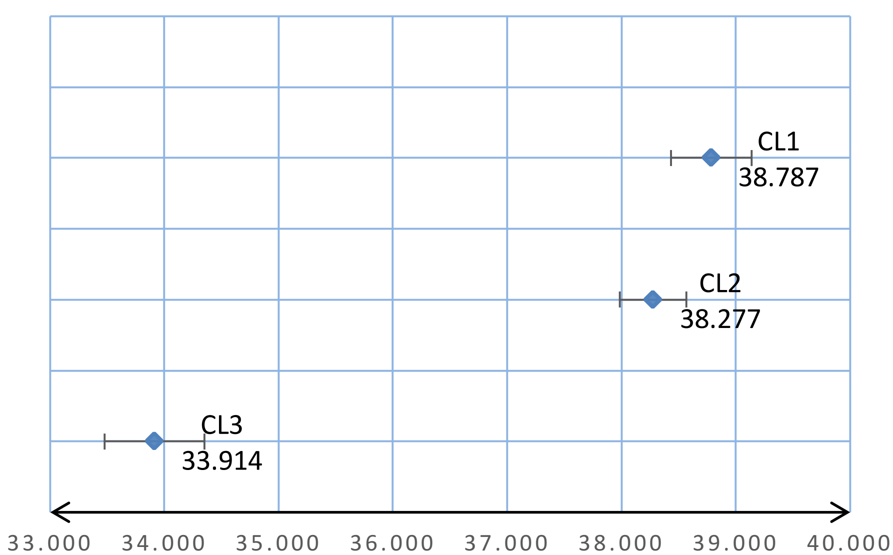


Figure 11 Estimated Means and CIs (X1RSCALK1-Fall 2010)

Table 18

Mean Estimates and Equality Tests of Means across Classes Using the BCH Procedure (Dependent Variable: X4RSCALK1-Spring 2012)

	Estimates	s.e.	MOE	LL	UL
CL1	71.668	0.252	0.494	71.174	72.162
CL2	71.543	0.210	0.412	71.131	71.955
CL3	62.708	0.373	0.731	61.977	63.439
Overall test	465.358	0.000			
CL1 vs. 2	0.116	0.051			
CL1 vs. 3	410.493	0.000			
CL2 vs. 3	363.724				

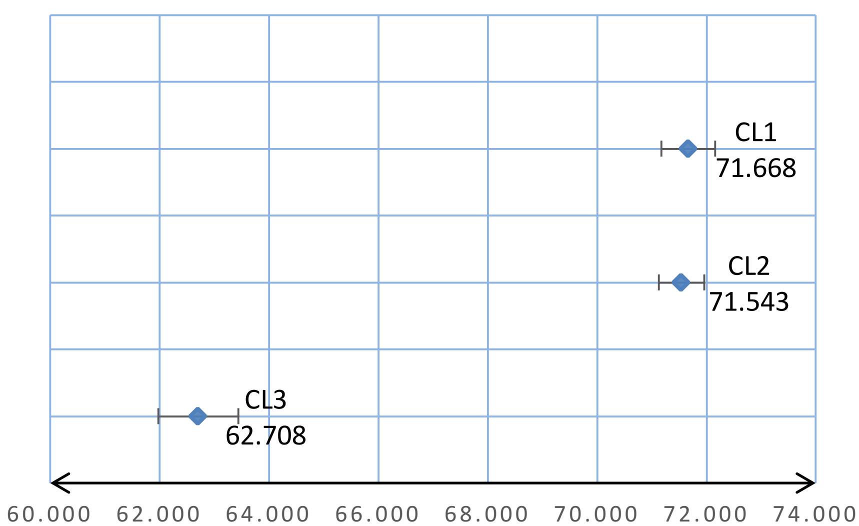


Figure 12 Estimated Means and CIs (X4RSCALK1-Spring 2012)

Model 1. The covariate, race, was added to the previous models to examine how the association between the early reading achievements (at the Fall Kindergarten and Spring Grade1) and racial group varies upon the identified latent classes. As shown in Table 19, the average reading score at the fall kindergarten decreases upon the exposure to the literacy environment, after controlling for the racial group: 39.535 for Class 1 (pro-reading1), 38.893 for Class 2 (pro-reading 2), and 34.393 for Class 3 (contra-reading). The magnitudes of each value were slightly inflated in comparison with the results of the model without the covariate. However, the changing patterns in the reading achievement in both models were similar to each other: there is no statistically significant difference between the two pro-reading groups, but significant discrepancy remains between those pro-reading groups and the contra reading groups.

As for the racial group comparison for each identified HLE profile, the reference category is Caucasian group, the three parameters for each dummy variable are the estimated differences in the reading scores from Caucasians. The positive values between the Asian group and the Caucasian group indicate that the children from Asian background outperform the Caucasian kids in the reading achievement in that particular HLE profile. The negative values indicate that Caucasian kids performed better than that comparison group. Hence, the Asian kids outperformed other racial groups in the reading achievement across all the HLE profiles. Moreover, the gaps shrink upon the degree of exposure to literacy activities. (6.612 for Class 1, 5.700 for Class 2, and 5.230 for Class 3). However, as shown in Table 19, the confidence intervals have highly overlapped across the three classes. Hence, the gaps were not significantly different from each other. As for the other racial group comparisons, in Class 1 (pro-reading 1), there was no significant difference between Hispanic kids and African-American kids in the early reading performance (2010 Kindergarten), but significant gaps exist compared with the Caucasian kids. Similar results were also found in Class 2. In contrast, there is no significant difference between Caucasian kids and African-American kids in the reading achievement in the Class 3 HLE profile (contra-reading), but the gap still exists between the Hispanic group and the Caucasian group.

Table 19

Parameter Estimates and Effect Sizes in Model 1 (Kindergarten Fall 2010)

Latent Class 1		Estimates	s.e.	MOE	LL	UP
X1RSCALK	D1 (African American/Caucasian)	-2.919	0.552	1.08192	-4.001	-1.837
	D2(Hispanic/Caucasian)	-3.662	0.465	0.9114	-4.573	-2.751
	D3(Asian/Caucasian)	6.612	1.323	2.59308	4.019	9.205
	Intercepts	39.535	0.218	0.42728	39.108	39.962
Latent Class 2						
X1RSCALK	D1 (African American/Caucasian)	-2.732	0.401	0.78596	-3.518	-1.946
	D2(Hispanic/Caucasian)	-3.500	0.357	0.69972	-4.200	-2.800
	D3(Asian/Caucasian)	5.700	0.801	1.56996	4.130	7.270
	Intercepts	38.893	0.194	0.38024	38.513	39.273
Latent Class3						
X1RSCALK	D1 (African American/Caucasian)	0.503	0.672	1.31712	-0.814	1.820
	D2(Hispanic/Caucasian)	-2.552	0.506	0.99176	-3.544	-1.560
	D3(Asian/Caucasian)	5.230	1.048	2.05408	3.176	7.284
	Intercepts	34.393	0.425	0.833	33.560	35.226
R-SQUARE						
Class1		Estimates	s.e.	MOE	LL	UP
X1RSCALK	RACE	0.051	0.010	0.0196	0.031	0.071
Class2						
X1RSCALK	RACE	0.059	0.009	0.0176	0.041	0.077
Class3						
X1RSCALK	RACE	0.062	0.013	0.0255	0.037	0.087

In terms of the effect size, the racial group only accounted for 5.1% - 6.2% of the variance in the early reading achievement, which is minimal but still statistically significant. This indicates that there are other potential predictors of reading achievement in addition to the participation in a racial group.

Model 2. The variable SES was added to the Latent Regression to investigate how the association between the demographic factors (i.e., SES and Race) and early reading achievement (2010 Spring Kindergarten) differs upon the different degree of exposure to the literacy interaction at home. The positive estimates of SES indicated the positive association of SES with the early reading achievement across the three HLE profiles. In addition, the two pro-reading groups were not significantly different from each other, but were significantly superior to the contra-reading group in the early reading achievement. These findings are similar to the results reported in the first two Models.

As for the racial membership comparisons for each HLE profile, at the average SES level, non-significant difference was found in these racial group comparisons: African-American group vs. Caucasian group in Class 1 and 2, Hispanic group vs. Caucasian group in Class 2 and 3. Furthermore, in the low literacy exposure HLE profile group (contra-reading), African American children had better performance than their Caucasian counterpart in the early reading achievement ($b = 1.477$, 95% CIs [0.193, 2.761]). Asian students still outperformed other groups in the early reading achievement across HLE profiles.

Table 20

Parameter Estimates and Effect Sizes in Model 2 (Kindergarten Fall 2010)

Latent Class 1		Estimates	s.e.	MOE	LL	UP
X1RSCALK	SES	4.237	0.253	0.49588	3.741	4.733
	D1 (African American/Caucasian)	-0.823	0.525	1.029	-1.852	0.206
	D2(Hispanic/Caucasian)	-1.518	0.446	0.87416	-2.392	-0.644
	D3(Asian/Caucasian)	4.720	1.302	2.55192	2.168	7.272
	Intercepts	38.477	0.202	0.39592	38.081	38.873
Latent Class 2						
X1RSCALK	SES	4.516	0.203	0.39788	4.118	4.914
	D1 (African American/Caucasian)	-0.231	0.393	0.77028	-1.001	0.539
	D2(Hispanic/Caucasian)	-0.672	0.355	0.6958	-1.368	0.024
	D3(Asian/Caucasian)	3.888	0.746	1.46216	2.426	5.350
	Intercepts	37.891	0.180	0.3528	37.538	38.244
Latent Class3						
X1RSCALK	SES	3.165	0.402	0.78792	2.377	3.953
	D1 (African American/Caucasian)	1.477	0.655	1.2838	0.193	2.761
	D2(Hispanic/Caucasian)	-0.784	0.521	1.02116	-1.805	0.237
	D3(Asian/Caucasian)	4.238	0.948	1.85808	2.380	6.096
	Intercepts	35.173	0.440	0.8624	34.311	36.035
R-SQUARE						
Class1		Estimates	s.e.	MOE	LL	UP
X1RSCALK	RACE+SES	0.153	0.014	0.0274	0.126	0.180
Class2						
X1RSCALK	RACE+SES	0.172	0.011	0.0216	0.150	0.194
Class3						
X1RSCALK	RACE+SES	0.109	0.018	0.0353	0.074	0.144

With regard to the effect size, the combination of Race and SES accounted for 10.9%-15.3% of variance in the early reading achievement for each HLE profile. The inclusion of SES in the model provides additional explanatory power, and SES has larger explanatory power compared to the racial group (see Table 20).

Results from the analytic model at the grade 1 spring 2012. I conducted three latent variable models with one distal outcome which is the reading achievement at the Grade 1 Spring 2012. The first two analytic models are similar to those of the Kindergarten 2011, with covariates of the racial group and SES. However, the third model has an additional covariate: the reading achievement of the kindergarten Fall 2010.

Model 1. Racial group is the only covariate in this model. Caucasian group is the reference category. Similar to findings of Model 1 at the kindergarten Fall 2010, Asian students outperformed all other racial groups in the reading achievement across the three HLE profiles. Hence, this reading achievement gap still remained at the Grade 1. Differently, the within-HLE profile racial group comparisons also showed statistically significant difference, but the cross-HLE profiles did not. The non-significant cross-HLE profile comparison was espoused by the highly overlapped confidence interval as shown in Table 21. In other words, there were no clear HLE profile-specific racial group differences in the reading achievement.

As for the explanatory power of the race, the variable race only explains 5.4%-6.5% of the variance in the reading achievement, which is similar to the results of Model 1 at the Kindergarten of Fall 2010.

Table 21

Parameter Estimates and Effect Sizes in Model 1 (Grade1 Spring 2012)

Latent Class 1		Estimates	s.e.	MOE	LL	UP
X4RSCALK	D1 (African American/Caucasian)	-6.849	0.883	1.731	-8.580	-5.118
	D2(Hispanic/Caucasian)	-5.393	0.729	1.429	-6.822	-3.964
	D3(Asian/Caucasian)	4.655	1.072	2.101	2.554	6.756
	Intercepts	73.250	0.299	0.586	72.664	73.836
Latent Class 2						
X4RSCALK	D1 (African American/Caucasian)	-4.949	0.653	1.280	-6.229	-3.669
	D2(Hispanic/Caucasian)	-6.331	0.551	1.080	-7.411	-5.251
	D3(Asian/Caucasian)	4.841	0.715	1.401	3.440	6.242
	Intercepts	72.599	0.261	0.512	72.087	73.111
Latent Class3						
X4RSCALK	D1 (African American/Caucasian)	-2.563	1.267	2.483	-5.046	-0.080
	D2(Hispanic/Caucasian)	-4.974	0.986	1.933	-6.907	-3.041
	D3(Asian/Caucasian)	4.483	1.362	2.670	1.813	7.153
	Intercepts	65.812	0.823	1.613	64.199	67.425
R-SQUARE						
Class1		Estimates	s.e.	MOE	LL	UP
X4RSCALK	RACE	0.054	0.009	0.018	0.036	0.072
Class2						
X4RSCALK	RACE	0.065	0.008	0.016	0.049	0.081
Class3						
X4RSCALK	RACE	0.060	0.014	0.027	0.033	0.087

Model 2. The covariates SES and race were added in Model 2. Asian students remained the top performers in the reading achievement. The significant racial gap between Caucasians and other racial groups were found in both pro-reading groups. In contrast, in the contra-reading group (low-exposure to home literacy), the racial gap between Caucasians and African-American/Hispanic groups were non-significant after controlling for SES. Finally, the SES and race also accounted for 13% and 15.8% of the variance in the reading achievement, but they were not significantly different from each other. The SES provided 7.6%-9.3% additional explanatory power. (see Table 23)

Model 3. The model 3 includes X1RSCALK1, SES, and race as covariates which is different from Models 1 and 2 at both Grades. The gap between Asian and Caucasian students in the reading achievement diminished into the non-significance across the HLE profiles, after controlling for SES and racial groups. Moreover, the reading gap between Hispanic and Caucasian remained non-significant ($b = -0.309$, 95% CI [-2.759, -0.523]), but other racial comparisons with Hispanic and African-American (Caucasian as the baseline) still showed the significant difference across the HLE profiles.

As for the explanatory power, the three predictors accounted for 41.5%-43.1% of the variance in the reading achievement at the Grade 1 Spring 2012. The inclusion of the earlier reading achievement provided significantly additional explanatory power in the model (27.3%-29%) in comparison with Model 2 (see Table 23).

In conclusion, the two pro-reading students outperformed their contra-reading counterparts in the reading achievement at both grade levels. Next, as for the overall racial comparison, Asian students outperformed other racial groups in the reaching achievement at both grade levels across the HLE profiles. However, the racial group gap between Caucasians and Hispanic/African-American children in both grade reading performance was not noticeable in the contra-reading group, after controlling for SES. Finally, the reading achievement at kindergarten had the strongest explanatory power in accounting for the variance in the reading achievement at Grade 1 in Fall 2012, in comparison with SES and racial groups across the HLE profiles.

Table 22

Parameter Estimates and Effect Sizes in Model 2 (Grade1 Spring 2012).

Latent Class 1		Estimates	s.e.	MOE	LL	UP
X4RSCALK	SES	4.841	0.341	0.668	4.173	5.509
	D1 (African American/Caucasian)	-4.498	0.857	1.680	-6.178	-2.818
	D2(Hispanic/Caucasian)	-2.872	0.704	1.380	-4.252	-1.492
	D3(Asian/Caucasian)	2.921	1.095	2.146	0.775	5.067
	Intercepts	71.965	0.308	0.604	71.361	72.569
Latent Class 2						
X4RSCALK	SES	5.533	0.277	0.543	4.990	6.076
	D1 (African American/Caucasian)	-1.685	0.631	1.237	-2.922	-0.448
	D2(Hispanic/Caucasian)	-2.578	0.557	1.092	-3.670	-1.486
	D3(Asian/Caucasian)	2.991	0.680	1.333	1.658	4.324
	Intercepts	71.193	0.265	0.519	70.674	71.712
Latent Class 3						
X4RSCALK	SES	5.635	0.632	1.23872	4.396	6.874
	D1 (African American/Caucasian)	-0.576	1.221	2.39316	-2.969	1.817
	D2(Hispanic/Caucasian)	-1.582	1.021	2.00116	-3.583	0.419
	D3(Asian/Caucasian)	3.403	1.25	2.45	0.953	5.853
	Intercepts	67.036	0.800	1.568	65.468	68.604
R-SQUARE						
Class 1		Estimates	s.e.	MOE	LL	UP
X4RSCALK	RACE+SES	0.130	0.013	0.0255	0.105	0.155
Class 2						
X4RSCALK	RACE+SES	0.158	0.010	0.0196	0.138	0.178
Class 3						
X4RSCALK	RACE+SES	0.136	0.018	0.0353	0.101	0.171

Table 23

Parameter Estimates and Effect Sizes in Model 3 (Grade1 Spring 2012)

Latent Class 1		Estimates	s.e.	MOE	LL	UP
X4RSCALK	SES	1.610	0.295	0.578	1.032	2.188
	X1RSCALK	0.747	0.022	0.043	0.704	0.790
	D1 (African American/Caucasian)	-4.039	0.699	1.370	-5.409	-2.669
	D2(Hispanic/Caucasian)	-1.600	0.579	1.135	-2.735	-0.465
	D3(Asian/Caucasian)	-0.589	0.920	1.803	-2.392	1.214
	Intercepts	43.175	0.945	1.852	41.323	45.027
Latent Class 2						
X4RSCALK	SES	2.194	0.240	0.470	1.724	2.664
	X1RSCALK	0.722	0.018	0.035	0.687	0.757
	D1 (African American/Caucasian)	-1.665	0.522	1.023	-2.688	-0.642
	D2(Hispanic/Caucasian)	-2.207	0.455	0.892	-3.099	-1.315
	D3(Asian/Caucasian)	0.036	0.596	1.168	-1.132	1.204
	Intercepts	43.832	0.756	1.482	42.350	45.314
Latent Class 3						
X4RSCALK	SES	2.051	0.511	1.002	1.049	3.053
	X1RSCALK	0.988	0.053	0.104	0.884	1.092
	D1 (African American/Caucasian)	-2.604	0.963	1.887	-4.491	-0.717
	D2(Hispanic/Caucasian)	-1.118	0.837	1.641	-2.759	0.523
	D3(Asian/Caucasian)	-0.309	1.067	2.091	-2.400	1.782
	Intercepts	32.456	1.950	3.822	28.634	36.278
R-SQUARE						
Class 1		Estimates	s.e.	MOE	LL	UP
X4RSCALK	RACE+SES +X1RSCALK	0.415	0.013	0.025	0.390	0.440
Class 2						
X4RSCALK	RACE+SES +X1RSCALK	0.431	0.011	0.022	0.409	0.453
Class 3						
X4RSCALK	RACE+SES +X1RSCALK	0.426	0.024	0.047	0.379	0.473

Study 2 Results

Descriptive statistics. Both parent-reported and teacher-reported socio-emotional competent items were used for the data analysis. The participants were tested at both grade levels using the same SEC items. Hence, there are four sets of SEC items to build the Latent Transitions Model, and two Latent Transition Models were built on the parent-reported and teacher-reported survey items respectively.

Model building. First, latent class models were estimated on both grade levels. Because the survey items are continuous variables, the latent class model would be estimated based on the different variance-covariance structures: class-independent and diagonal, class-dependent and diagonal, class-independent and unrestricted, and class-dependent and unrestricted. Table 24 presents the maximum number of latent classes that can be successfully identified for both teacher-reported and parent-reported survey times at both grade levels.

Table 24

The Maximum Number of Latent Classes to Obtain the Identifiable Model in Latent Profile Model on Time 1 and Time2 for both Parents and Teacher SEC Measures.

	Parent T1	Parent T2	Teacher T1	Teacher T2
Class-indep, diag	6	6	6	6
Class-dep,diag	2	2	3	3
Class-indep, full	6	6	6	6
Class-dep,full	2	2	2	2

Model selection. Tables 25 to 28 display the model fit indices for each variance-covariance matrix. The model selection is based on the values of BIC, AIC, CAIC, and AWE. The model with smaller values is preferred. As for parent-reported items, the 2-class model with the class-dependent and unrestricted variance-covariance matrix is selected based on the smaller model fit indices in comparison with the other 2-class models. Another rationale for the selection is based on the latent class separation. The 3-class solution cannot give the clear class separation for all conditions. Hence, the 2-class dependent and unrestricted models were selected as the final models for the latent transition model.

Table 25

Model Fits of Latent Profile Analysis at the Time 1 on Parent-Reported SEC Measures

Structure	# of class	n	LL	BIC	AIC	CAIC	AWE
Class-indep, diag	1	10	-49563	99221	99146	99231	99346
	2	16	-46602	93357	93237	93373	98468
	3	22	-45352	90914	90749	90936	97947
	4	28	-44455	89177	88967	89205	97561
	5	34	-43871	88065	87810	88099	97204
	6	40	-43554	87487	87187	87527	97923
Class-dep, diag	1	10	-49563	99221	99146	99231	99346
	2	21	-44474	89148	88990	89169	93966
Class-indep, full	1	20	-45050	90289	90139	90309	90539
	2	26	-44087	88422	88227	88448	93187
	3	32	-43587	87477	87237	87509	92452
	4	38	-43345	87051	86766	87089	94130
	5	44	-42793	86005	85675	86049	94541
	6	50	-42373	85221	84846	85271	89086
Class-dep, full	1	20	-45050	90289	90139	90309	90539
	2	41	-30368	61126	60818	61167	62145

Table 26

Model Fits of Latent Profile Analysis at the Time 2 on Parent-Reported SEC Measures

Structure	# of class	n	LL	BIC	AIC	CAIC	AWE
Class-indep, diag	1	10	-36228	72551	72476	72561	72676
	2	16	-33730	67613	67493	67629	75435
	3	22	-32844	65897	65732	65919	77280
	4	28	-32101	64469	64259	64497	77971
	5	34	-31662	63648	63393	63682	77916
	6	40	-31278	62935	62635	62975	78480
Class-dep, diag	1	10	-36228	72551	72476	72561	72676
	2	21	-32106	64411	64253	64432	72256
Class-indep, full	1	20	-32412	65014	64864	65034	65264
	2	26	-31788	63822	63627	63848	71202
	3	32	-31107	62518	62278	62550	71994
	4	38	-30963	62287	62002	62325	74152
	5	44	-30427	61273	60943	61317	71071
	6	50	-30427	61329	60954	61379	80880
Class-dep, full	1	20	-32412	65014	64864	65034	65264
	2	41	-21235	42860	42553	42901	48072

Table 27

Model Fits of Latent Profile Analysis at the Time 1 on Teacher-Reported SEC Measures

Structure	# of class	n	LL	BIC	AIC	CAIC	AWE
Class-indep, diag	1	10	-53936	107967	107892	107977	108092
	2	16	-43397	86946	86826	86962	90987
	3	22	-39384	78977	78812	78999	86031
	4	28	-37807	75880	75670	75908	84743
	5	34	-37278	74880	74625	74914	84597
	6	40	-36797	73973	73673	74013	87031
Class-dep, diag	1	10	-53936	107967	107892	107977	108092
	2	21	-41296	82791	82634	82812	87381
	3	32	-36865	74033	73793	74065	80732
Class-indep, full	1	20	-38465	77119	76969	77139	77369
	2	26	-37149	74544	74349	74570	77625
	3	32	-36162	72628	72388	72660	77662
	4	38	-35780	71922	71637	71960	80184
	5	44	-35425	71267	70937	71311	80607
	6	50	-34741	69956	69581	70006	78239
Class-dep, full	1	20	-38465	77119	76969	77139	77369
	2	41	-34705	69801	69493	69842	74644

Table 28

Model Fits of Latent Profile Analysis at the Time 2 on Teacher-Reported SEC Measures

Structure	# of class	n	LL	BIC	AIC	CAIC	AWE
Class-indep, diag	1	10	-46409	92914	92839	92924	93039
	2	16	-36479	73110	72990	73126	79050
	3	22	-33082	66373	66208	66395	76425
	4	28	-32048	64363	64153	64391	77435
	5	34	-31604	63531	63276	63565	79019
	6	40	-31199	62777	62477	62817	78979
Class-dep, diag	1	10	-46409	92914	92839	92924	93039
	2	21	-34653	69505	69347	69526	76217
	3	32	-30676	61655	61415	61687	71916
Class-indep, full	1	20	-32762	65714	65564	65734	65964
	2	26	-31778	63804	63609	63830	68161
	3	32	-31264	62833	62593	62865	68988
	4	38	-30805	61972	61687	62010	72982
	5	44	-30516	61451	61121	61495	75291
	6	50	-30428	61331	60956	61381	78165
Class-dep, full	1	20	-32762	65714	65564	65734	65964
	2	41	-29612	59613	59305	59654	65944

What is the latent profile of SEC, and its variation in membership on covariates including SES and gender? The latent class models on both grade levels were conducted on the five parent-reported SEC measures (see Table 29 and 30). Figure 13 and 14 present the averages of items for each latent class. The visualized plots indicate the measurement invariance across the two time points. The interpretation of the identified latent class is based on the five estimated average of SEC items within each latent class. These five items are Approaches to learning, self-control, social-interaction, feeling sad and loneliness, and impulsive/overactive behaviors. The higher scores indicate the children exhibited the behavior measured by the scale more often. Results of the LCA on parent-reported SEC scales identified two latent classes. The first latent class is characterized by the lower score on approaches to learning, self-control, social-interaction, but higher scores on impulsive/overactive behavior and feeling sad and loneliness. The second class bears the opposite features. Hence, the children belonging to the second class exhibited those behaviors more often indicative of the positive learning behavior, sell-control, better interpersonal skills, but less frequency in overactive behaviors and feeling sad or lonely. In contrast, the students classified into the first class exhibited less often in the three positive socio-emotional behaviors, but more often in the two negative socio-emotional behaviors. Briefly, the identified latent class 1 is named the positive SEC group, and the latent class 2 is named the negative SEC group.

The measurement invariance is assumed between the two time points, as espoused by the similar Latent Classification patterns at the Grade 1. Hence, the identified latent classes at the second time points bear the same features: the positive SEC group (class 2), and the negative SEC group (class 1).

Table 29

Estimated Means of Latent Profile Analysis on Parent-Reported Scales for Each Latent Class at Time 1

Class 1					
	Estimates	s.e.	MOE	LL	UL
Class Size	0.655	0.007			
X1PRNAPP	3.053	0.006	0.012	3.041	3.065
X1PRNCON	2.803	0.007	0.013	2.790	2.816
X1PRNSOC	3.191	0.008	0.015	3.176	3.206
X1PRNSAD	1.587	0.005	0.009	1.577	1.596
X1PRNIMP	2.107	0.009	0.017	2.090	2.123
Class 2					
Class Size	0.345	0.007			
X1PRNAPP	3.417	0.008	0.015	3.402	3.433
X1PRNCON	3.054	0.009	0.018	3.037	3.072
X1PRNSOC	3.906	0.003	0.006	3.900	3.912
X1PRNSAD	1.297	0.005	0.010	1.287	1.306
X1PRNIMP	1.920	0.012	0.024	1.895	1.944

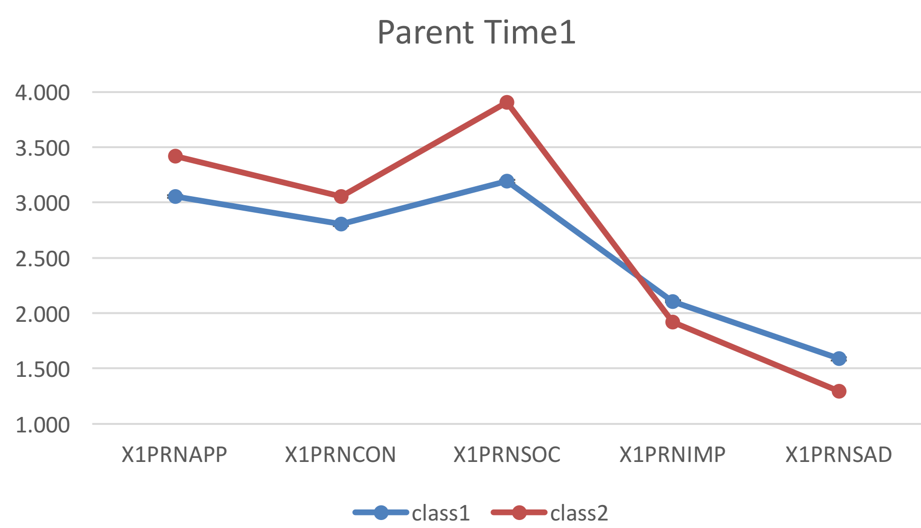


Figure 13 Plot of LPA Estimated Means of Parent-Reported Scales for Each Latent Class at Time1

Table 30

Estimated Means of Latent Profile Analysis on Parent-Reported Scales for Each Latent Class at Time 2

Class 1	Estimates	s.e.	MOE	LL	UP
Class Size	0.664	0.005			
X4PRNAPP	2.963	0.006	0.012	2.951	2.974
X4PRNCON	2.951	0.006	0.012	2.939	2.963
X4PRNSOC	3.184	0.006	0.011	3.172	3.195
X4PRNSAD	1.516	0.005	0.010	1.506	1.526
X4PRNIMP	1.907	0.008	0.016	1.891	1.923
Class 2					
Class Size	0.336	0.005			
X4PRNAPP	3.359	0.007	0.014308	3.345	3.374
X4PRNCON	3.149	0.008	0.015876	3.134	3.165
X4PRNSOC	4.000	0.000	0.000196	4.000	4.000
X4PRNSAD	1.369	0.006	0.010976	1.358	1.380
X4PRNIMP	1.812	0.011	0.021756	1.790	1.834

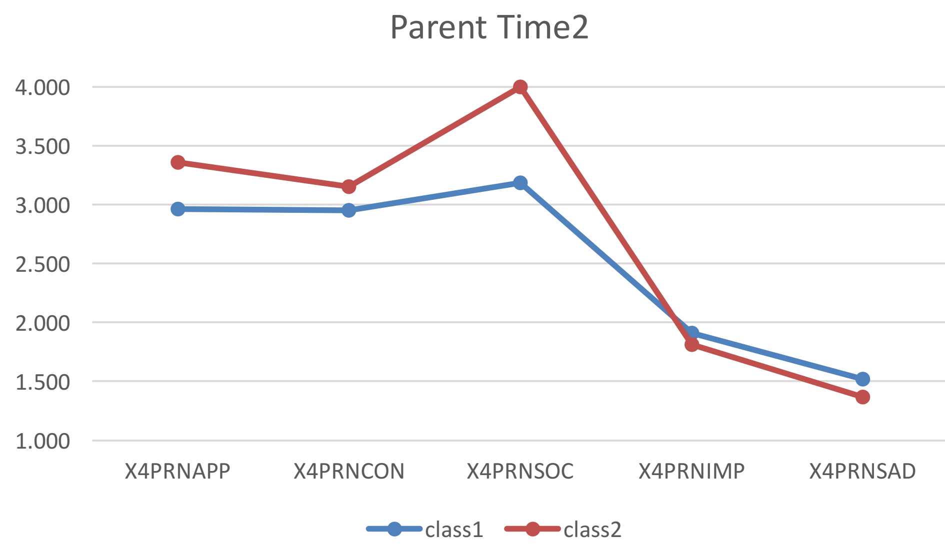


Figure 14 Plot of LPA Estimated Means of Parent-Reported Scales for Each Latent Class at Time 2

Are the SEC profiles stable over time? What is the long-term effect of SEC on the later reading comprehension? To test the stability of SEC over the time, Latent Transition Analysis was conducted between the kindergarten and grade 1 with restriction on the same item estimated mean, and variance-covariance matrices across the two time points. Table 31 presents the mover-and-stayer patterns represented by the transition probabilities.

As shown in Table 31, the probability of staying in the negative SEC group is 0.862, and the probability of moving from the negative to the positive SEC group is 0.138. The likelihood of staying in the positive SEC group is 0.756, while the likelihood of moving from the positive to the negative group is 0.244.

Table 31

Latent Transition Probabilities from Kindergarten to Grade 1 Based on Parent-Reported Scales

State 1	State 2	Transition Patterns	Prob	s.e.	MOE	LL	UL
1	1	1	0.862	0.007	0.013	0.849	0.875
1	2	2	0.138	0.007	0.013	0.125	0.151
2	1	3	0.244	0.011	0.022	0.222	0.266
2	2	4	0.756	0.011	0.022	0.734	0.778

Note: State 1 is the negative SEC behavior group, and State 2 is the positive SEC behavior group; Transition pattern: 1 = staying in the negative behavior group; 2 = moving from negative to positive SEC behavior group, 3 = moving from positive to negative SEC behavior group, 4 = staying in the positive SEC behavior group

When the transitional probabilities are split by the gender, male children are more likely to stay in the negative SEC group than the female children: male (0.892) vs. female (0.827). In contrast, the male children seem to have the lower likelihood of staying in the positive SEC group than the female children: male (0.744) vs. female (0.769). However, the difference is not statistically significant, because of the overlapped confidence intervals for the two estimated probabilities (see Table 32).

As for the probabilities for the movers, although the male children have the higher estimated probability of moving from the positive to the negative SEC group, there is no indication of a significant difference between the two estimates based on the overlapped confidence interval. However, female children have significantly higher likelihood of moving from the negative to the positive SEC group than the male children: male (0.108, 95% CIs [0.092, 0124]) vs. female (0.173, 95% CIs [0.153, 0.192]).

To examine how the latent transition probabilities vary upon the SES and gender, the SES was added to the LTA model. Table 33 showed that the higher SES children tend to be more likely to stay in the positive SEC group and less likely to stay in the negative SEC group. Moreover, the children from the higher SES tend to be more likely to move from the negative SEC to the positive SEC, and less likely to move from the positive SEC to the negative SEC group for both genders.

Table 32

Latent Transition Probabilities from Kindergarten to Grade 1 Based on Parent-Reported Scale by Gender

Gender	State 1	State 2	Prob	s.e.	MOE	LL	UL
Male	1	1	0.892	0.008	0.016	0.876	0.908
	1	2	0.108	0.008	0.016	0.092	0.124
	2	1	0.256	0.017	0.034	0.222	0.290
	2	2	0.744	0.017	0.034	0.710	0.778
Female	1	1	0.827	0.010	0.020	0.808	0.847
	1	2	0.173	0.010	0.020	0.153	0.192
	2	1	0.231	0.014	0.028	0.203	0.259
	2	2	0.769	0.014	0.028	0.741	0.797

Table 33

Latent Transition Probabilities from Kindergarten to Grade 1 Based on Parent-Reported Scale by Gender and SES

SES	Gender	State 1	State 2	Prob	s.e.	MOE	LL	UL
-2.330	Male	1	1	0.944	0.009	0.017	0.927	0.961
		1	2	0.056	0.009	0.017	0.039	0.073
		2	1	0.397	0.051	0.100	0.296	0.497
		2	2	0.604	0.051	0.100	0.503	0.704
	Female	1	1	0.907	0.014	0.026	0.880	0.933
		1	2	0.093	0.014	0.026	0.067	0.120
		2	1	0.363	0.048	0.093	0.269	0.456
		2	2	0.638	0.048	0.093	0.544	0.731
0.000	Male	1	1	0.891	0.008	0.016	0.875	0.907
		1	2	0.109	0.008	0.016	0.093	0.125
		2	1	0.263	0.018	0.035	0.228	0.297
		2	2	0.737	0.018	0.035	0.703	0.772
	Female	1	1	0.825	0.010	0.020	0.805	0.845
		1	2	0.175	0.010	0.020	0.155	0.195
		2	1	0.236	0.015	0.028	0.207	0.264
		2	2	0.764	0.015	0.028	0.736	0.793
2.600	Male	1	1	0.784	0.028	0.055	0.729	0.840
		1	2	0.216	0.028	0.055	0.160	0.271
		2	1	0.153	0.027	0.053	0.099	0.206
		2	2	0.847	0.027	0.053	0.794	0.901
	Female	1	1	0.677	0.035	0.069	0.608	0.747
		1	2	0.323	0.035	0.069	0.253	0.392
		2	1	0.135	0.025	0.048	0.087	0.183
		2	2	0.865	0.025	0.048	0.817	0.913

Results of validation based on teacher-reported scale. The series of Latent Transition Analyses were also computed for the teacher-reported items for the validation purpose. The analysis includes five items for each condition: approaches to learning, self-control, externalizing behaviors problem (e.g., physical aggression, breaking rules, cheating, stealing, and vandalism), and internalizing problem (e.g., depression, withdrawal, anxiety, and loneliness). Higher score in the item that exhibited the behavior was represented by the scale more often. For example, the higher score in the impersonal skills infer that the children exhibited positive social interaction behaviors more often, and the higher scores in the internalizing problem behaviors mean the children suffered depression, demonstrated social withdrawal, anxiety, and loneliness more often. Similar to the latent class solution based on the parent-reported socio-emotional competence scale, the teacher-reported scales also include three positive behavior items, and two negative behavior items. Hence, the teacher-reported and the parent-reported social skills cover similar constructs.

Likewise, the Latent Class Analysis identified two latent class solutions: positive SEC group and negative SEC group at both grade levels, as well as the similar latent class prevalence structure as shown in Figures 15 to 16, and Tables 34 to 37. The follow-up Latent Transition Analysis also assumed the measurement invariance across the two time points.

Table 34

Estimated Means of Latent Profile Analysis Based on Teacher-Reported Scales for Each Latent Class on Time 1

	Estimates	s.e.	MOE	LL	UL
Class 1					
Class Size	0.541	0.010			
X1TCHAPP	2.696	0.011	0.021	2.676	2.717
X1TCHCON	2.840	0.010	0.020	2.820	2.859
X1TCHPER	2.745	0.010	0.019	2.726	2.765
X1TCHEXT	1.870	0.010	0.020	1.850	1.890
X1TCHINT	1.714	0.011	0.021	1.694	1.735
Class 2					
Class Size	0.459	0.010			
X1TCHAPP	3.324	0.013	0.025	3.299	3.349
X1TCHCON	3.407	0.012	0.024	3.383	3.431
X1TCHPER	3.335	0.012	0.023	3.311	3.358
X1TCHEXT	1.278	0.009	0.018	1.260	1.296
X1TCHINT	1.156	0.003	0.007	1.149	1.162

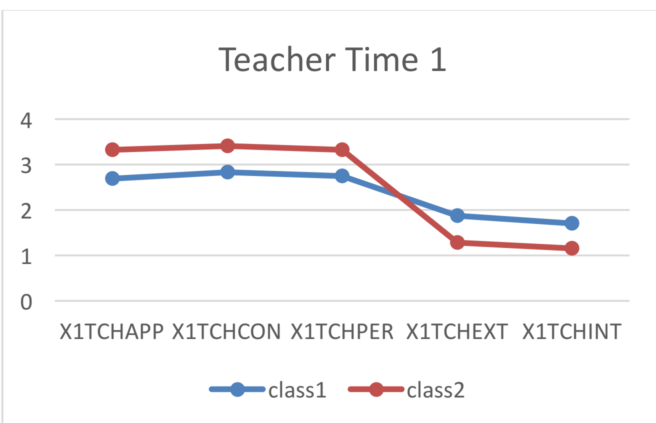


Figure 15 Plot of LPA Estimated Means of Teacher-Reported Scales for Each Latent Class at Time 1

Table 35

Estimated Means of Latent Profile Analysis Based on Teacher-Reported Scales for Each Latent Class on Time 2

	Estimates	s.e.	MOE	LL	UP
Class 1					
Class Size	0.764	0.007			
X4TCHAPP	2.873	0.009	0.018	2.855	2.891
X4TCHCON	3.048	0.008	0.015	3.033	3.064
X4TCHPER	2.960	0.009	0.017	2.943	2.977
X4TCHEXT	1.855	0.008	0.016	1.839	1.870
X4TCHINT	1.632	0.007			
Class 2					
Class Size	0.236	0.007			
X4TCHAPP	3.829	0.008	0.015	3.814	3.844
X4TCHCON	3.799	0.008	0.016	3.783	3.815
X4TCHPER	3.790	0.008	0.016	3.773	3.806
X4TCHEXT	1.270	0.007	0.013	1.256	1.283
X4TCHINT	1.220	0.006	0.011	1.209	1.231

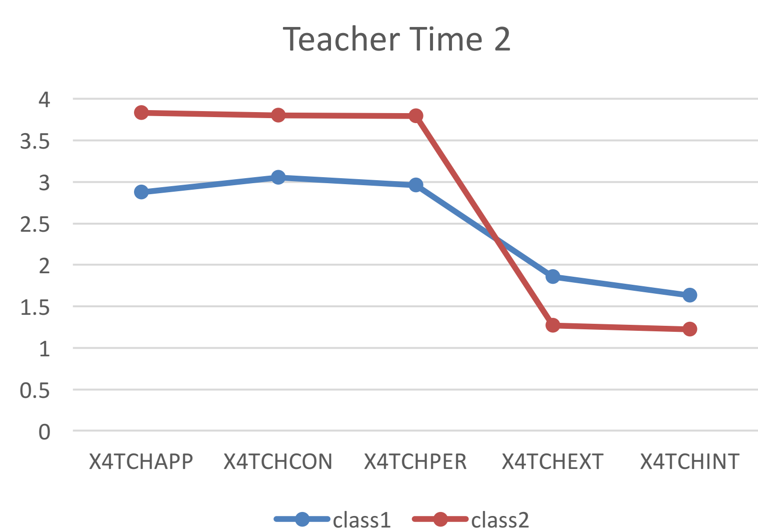


Figure 16 Plot of LPA Estimated Means of Parent-Reported Scales for Each Latent Class at Time 2

Table 36 presents the results of the Latent Transition Analysis indicating that the probabilities of staying in the negative SEC group is 0.802, and in the positive group is 0.667. The likelihood of moving from the negative to the positive SEC group is 0.199, and from the positive to negative SEC group, the transition probability is 0.333. The analysis split by gender (see Table 37) indicated that male children have a higher probability of staying in the negative SEC group but lower probability of staying in the positive SEC group in comparison with their female counterparts. Moreover, the male students are more likely to move from the positive SEC group to the negative SEC group: male (0.354, 95% CIs [0.320, 0.388]) vs. female (0.300, 95% CIs [0.274, 0.325]). In contrast, female students have significantly higher likelihood of moving from the negative SEC group to the positive SEC group: female (0.281, 95% CIs [0.256, 0.306]) vs. male (0.133, 95% CIs [0.116, 0.150]).

Table 36

Latent Transition Probabilities from Kindergarten to Grade 1 Based on Teacher-Reported Scale

State 1	State 2	Transition Patterns	Prob	s.e.	MOE	LL	UL
1	1	1	0.802	0.008	0.015	0.786	0.817
1	2	2	0.199	0.008	0.015	0.183	0.214
2	1	3	0.333	0.011	0.022	0.311	0.356
2	2	4	0.667	0.011	0.022	0.644	0.689

Table 37

Latent Transition Probabilities from Kindergarten to Grade 1 Based on Teacher-Reported Scale by Gender

Gender	State 1	State 2	Prob	s.e.	MOE	LL	UP
Male	1	1	0.867	0.009	0.017	0.850	0.884
	1	2	0.133	0.009	0.017	0.116	0.150
	2	1	0.354	0.017	0.034	0.320	0.388
	2	2	0.646	0.017	0.034	0.612	0.680
Female	1	1	0.719	0.013	0.025	0.694	0.744
	1	2	0.281	0.013	0.025	0.256	0.306
	2	1	0.300	0.013	0.025	0.274	0.325
	2	2	0.700	0.013	0.025	0.675	0.726

Table 38 presents the transition probabilities for each gender and three SES levels. The results demonstrated similar patterns to the findings from the analysis based on the parent-reported scales. Specifically, children with higher SES levels tended to stay in the positive SEC group and less probability of staying in the negative SEC group compared with children from the lower SES families. Within each SES level, female students still demonstrated higher likelihood of exhibiting positive behaviors and transitioning from the negative SEC group to the positive SEC group. Briefly, the findings based on the teacher-reported scales support and validate the conclusion of the analysis based on the parent-reported scales.

Table 38

Latent Transition Probabilities from Kindergarten to Grade 1 Based on Teacher-Reported Scale by Gender and SES

X12SESL	X_CHSEX_R	Class1	Class2	Prob	s.e.	MOE	LL	UL
-2.330	Male	1	1	0.953	0.008	0.015	0.938	0.968
		1	2	0.047	0.008	0.015	0.032	0.062
		2	1	0.497	0.041	0.080	0.417	0.577
		2	2	0.503	0.041	0.080	0.423	0.583
	Female	1	1	0.886	0.016	0.031	0.854	0.917
		1	2	0.114	0.016	0.031	0.083	0.146
		2	1	0.425	0.037	0.072	0.352	0.497
		2	2	0.575	0.037	0.072	0.503	0.648
0.000	Male	1	1	0.866	0.009	0.018	0.848	0.883
		1	2	0.134	0.009	0.018	0.117	0.152
		2	1	0.358	0.018	0.034	0.324	0.393
		2	2	0.642	0.018	0.034	0.607	0.676
	Female	1	1	0.711	0.013	0.026	0.685	0.737
		1	2	0.289	0.013	0.026	0.263	0.315
		2	1	0.294	0.013	0.026	0.269	0.320
		2	2	0.706	0.013	0.026	0.680	0.731
2.600	Male	1	1	0.642	0.038	0.076	0.566	0.718
		1	2	0.358	0.038	0.076	0.282	0.434
		2	1	0.228	0.028	0.055	0.172	0.283
		2	2	0.772	0.028	0.055	0.717	0.828
	Female	1	1	0.406	0.041	0.081	0.325	0.486
		1	2	0.594	0.041	0.081	0.514	0.675
		2	1	0.181	0.024	0.047	0.134	0.227
		2	2	0.819	0.024	0.047	0.773	0.866

To assess the long-term reading outcomes of the change process of the socio-emotional competence, the Grade 1 reading achievement as the distal outcome was added to the latent transition model with the Kindergarten reading achievement, SES, and gender as covariates. Then, a Wald test of the mean contrasts was conducted to determine whether there was a significant difference in the reading achievement across classes at time 2. The analysis was also conducted for both parent-reported and teacher-reported scales. As shown in Table 39, the analysis based on the parent-reported scales ($M=76.44$, $SE=0.198$) indicated that students in the positive SEC group significantly outperformed their counterparts in the negative SEC group ($M=58.318$, $SE=0.508$), after controlling for the earlier reading achievement, SES and gender (Chi-squared = 1081.551, $df=1$, $p<0.001$). The results were also espoused by the analysis based on the teacher-reported scales and revealed the same patterns (Chi-squared=1230.156, $df=1$, $p<.0001$).

Table 39

The Long-Term Effects of SEC on Reading Achievement at Grade 1 Based on Parent-Reported and Teacher-Reported Scales

Parent State 2	Estimates	s.e.	MOE	LL	UP
1	58.318	0.508	0.996	57.322	59.314
2	76.440	0.198	0.388	76.052	76.828
Wald Testing					
	Value	1081.551			
	df	1			
	p-value	<0.0001			
Teacher State 2					
1	62.513	0.421	0.825	62.092	63.338
2	77.252	0.167	0.327	77.085	77.579
Wald Testing					
	Value	1230.156			
	df	1			
	p-value	<.0001			

Study 3 Results

Sample size and missing value. The final data set for the Decision Tree Model is the merged data by combining the results dataset from the Latent Transition Model. The sample size for the Decision Tree Model is 12,730.

The Grade 1 reading achievement score was dichotomized on the bottom 10th percentile value: 51.04 to create the indicator for at-reader reader: ATRISK. If the reading achievement score was below 51.04, then ATRISK = 1, otherwise, ATRISK = 0. Finally, there are 1,086 children belonging to at-risk reader group, and 9,762 children in non-at-risk reader group. The average reading achievement scores are 73.03 (SD=10.32) for non-at-risk readers, and 44.304 (SD = 5.38) for at-risk readers. The Decision Tree analysis was conducted via SAS Enterprise Miner. Figure 17 display the Decision Tree panel in SAS Enterprise Miner interface.

Variable inclusion. The target variable in the Classification Model is the binary indicator for at-risk reader (ATRISK). The input variables include Grade 1 Fall 2010 (RDNG1), Transition Patterns based on the teacher-reported scale (TTRN), Transition patterns based on the parent-reported scale (PTRN), the racial group (X_RACE), gender (X_CHSEX_R), and the identified 3-Class HLE profiles. The missing values are included into the modeling. The Largest Brank option was selected for handling the missing values: all of the observations with the missing values for the splitting rule would be placed in the branch with the largest number of training observations.

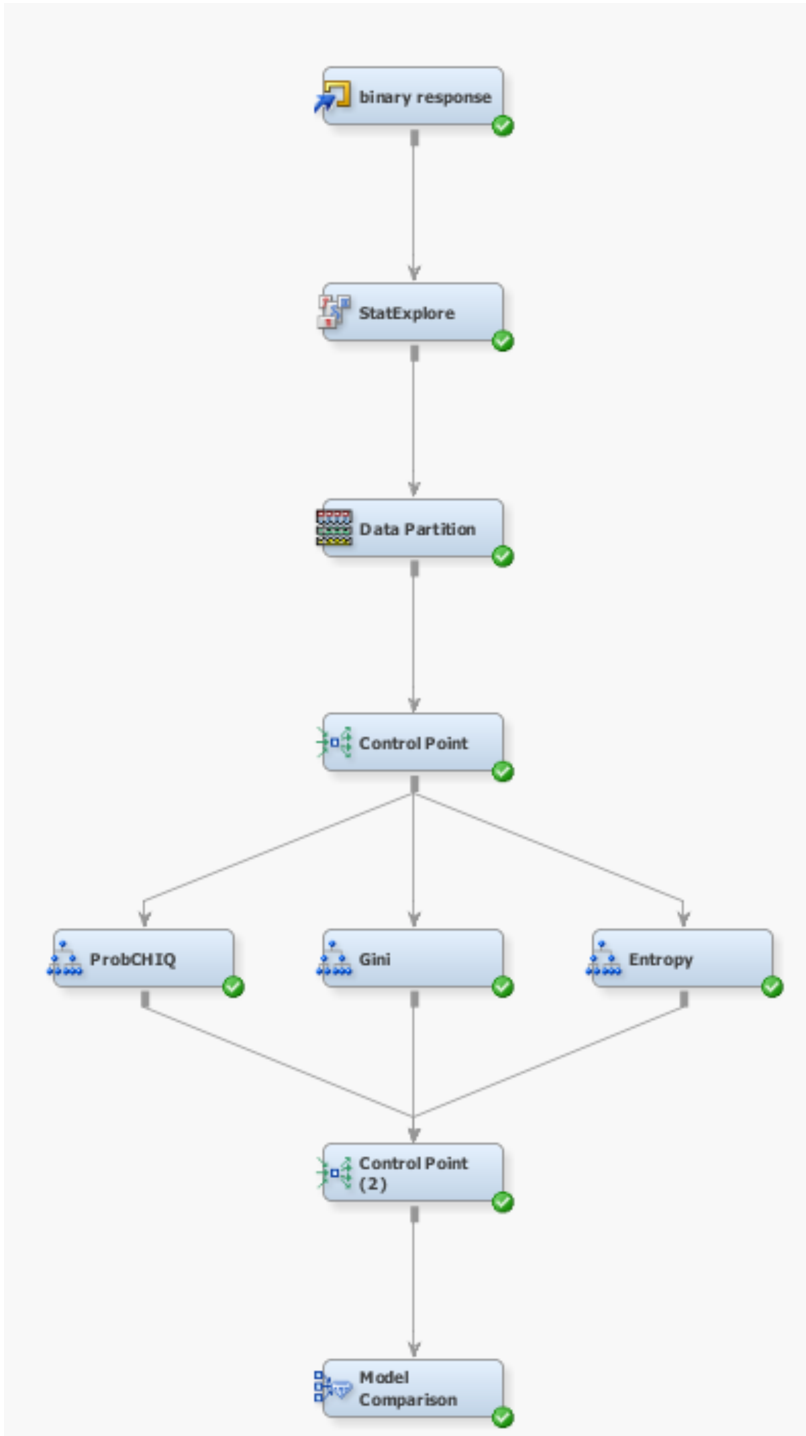


Figure 17 Decision Tree Model Building in SAS Enterprise Miner

The sample was also partitioned into the training data (70%) and validation (30%) for model building and selection. In the Classification Tree Model, three splitting rules were used for model building: ProbCHIQ, Gini, and Entropy. The results based on the validation were used for model selection. The selection criterion is the misclassification rate based on the validation data results. The model with the smaller misclassification rate was preferred. The model with Gini split rule had the smallest value (0.084) in comparison with the model with ProbCHISQ (0.086) and the model with Entropy (0.086).

The variable importance in Table 40 presents the relative importance of the input variables in the Classification Tree Model. This table shows which variables are important for the identifying the at-risk readers. Three indices were used: importance (training), validation importance, and ratio of validation to training variable importance. Higher scores in the importance and validation importance statistics indicated that variables have more contribution in splitting, and classification. The ratio close 1 measures the consistency in identifying the important predictors between training and validation data.

As shown in Table 40, the reading achievement at the Kindergarten Fall 2010 is the most important predictor in identifying the at-risk readers. Other relatively important predictors are the teacher-reported SEC change patterns (TTRN), SES, gender (X_CHSEX_R), and parent-reported SEC change patterns (PTRN). The racial group variable indicated inconsistency in training and validation dataset. The identified home literacy environment profiles were not included in the tree growing process, indicative of its high correlation with other variables.

Table 40

The Variable Importance Based on the Gini Splitting Rules

Variable name	Importance	Validation Importance	Ratio of importance to Validation importance
RDNG1	1	1.0000	1.0000
TTRN	0.2679	0.1991	0.7431
SES	0.1605	0.2004	1.2487
X_RACE	0.1091	0.0000	0.0000
X_CHSEX_R	0.1072	0.1230	1.1472
PRTN	0.1002	0.0600	0.5994
CLS	0.0000	0.0000	.

To be more specific, as shown in Figure 18, the first splitting on the score of 28.10 of the kindergarten reading achievement. If the early reading score is lower than 28.10, the students in that group have the probability of 37.89% to classified as at risk-readers. In contrast, if the early reading score is higher than 28.104, then the probability is only 4.52%. The highest probability of being classified into at-risk readers was found by the splitting on the parent-reported change behavior. The value is 62.39% in the training set, and 54.72% in the validation set. Figure 18 also presents the tree results, which can be used to profile that particular group. For example, students that tend to have a higher likelihood of being in at-risk readers are: male students, the kindergarten reading score less than 28.10 (percentile), exhibiting negative or transitioning to the negative SEC behavior based on the teacher-reported scale and initially exhibiting negative SEC groups (though transitioning to the positive SEC group) based on the parent-reported scales, coming from the low SES family (less than -0.745) and Hispanic families. If the students were observed to have the behaviors described above, they might need to be further diagnosed by a reading specialist and a school psychologist for verification. The teacher can give special attention, extra instruction, and immediate intervention to that particular group.

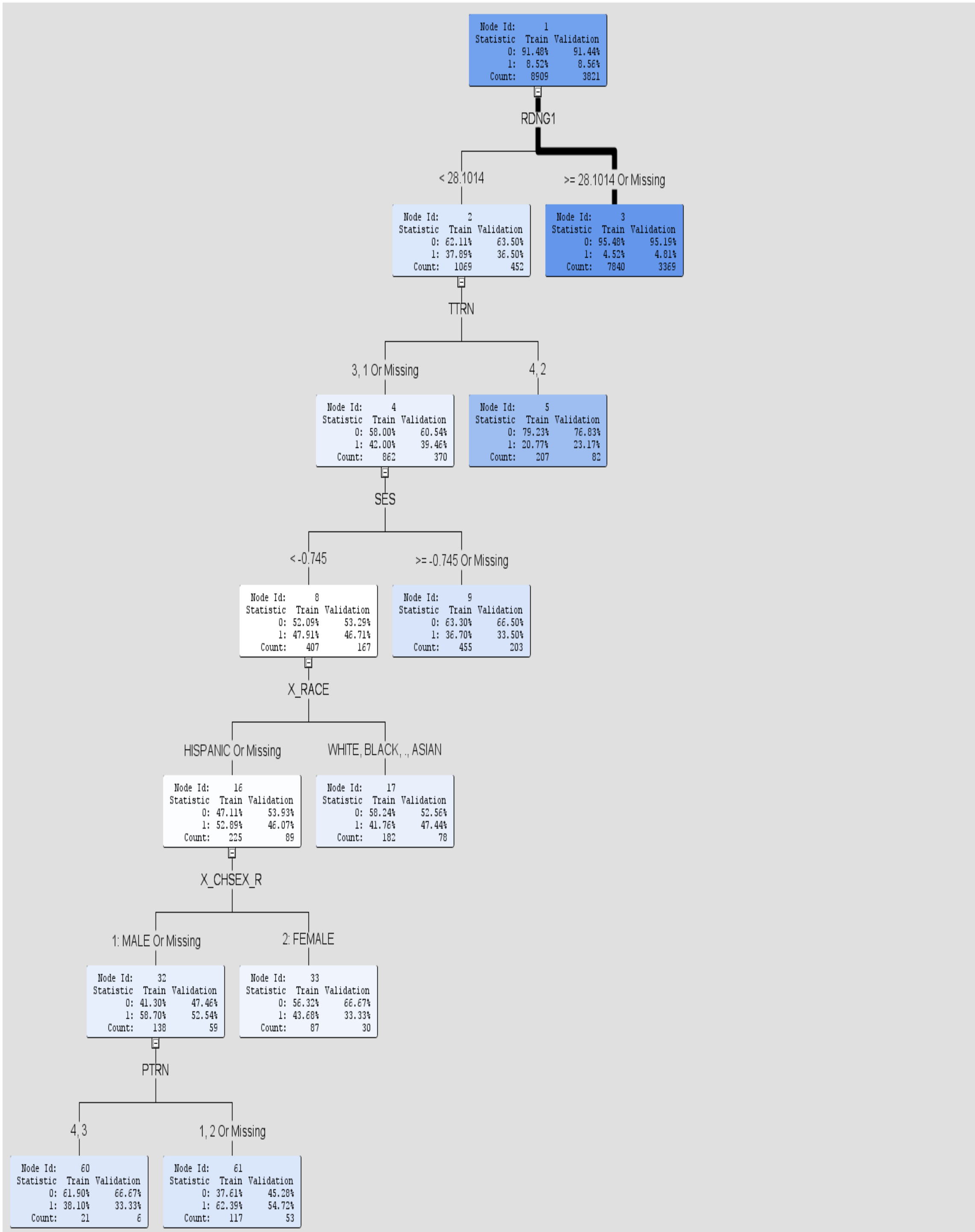


Figure 18 Decision Tree Diagram

CHAPTER V

CONCLUSIONS AND RECOMMEDATIONS

The purposes of this study were to (1) identify the underlying types of home literacy environment (HLE) and its effect on the early reading achievement, (2) examine the types of socio-emotional competence (SEC) and its long-term effect on the later reading achievement, and (3) investigate the contribution of the HLE and SEC to identify the at-risk readers. The HLE and SEC play essential roles in the reading development. However, the construct of the HLE is always simply defined by a single measure, and lacks a consensus definition and the conceptualization of the SEC. In addition to the measurement and conceptualization issues, few studies have examined the stability of SEC over time and its long-term effect on the later reading achievement. Moreover, there is also a lack of investigation into the complex interactions among HLE, SEC, and demographic factors on reading achievement.

This study utilized Componential Model of Reading (Joshi & Aaron, 2000, 2012; Joshi, Tao, Aaron, & Quiroz, 2012) and Ecological System Theory (Bronfenbrenner, 1986, 1994) as theoretical frameworks. Three interconnected studies were conducted to examine the effects of the interaction between the ecological, cognitive, and psychological factors on identifying the at-risk readers.

Study1 attempted to address three research questions: (1) What are the HLE profiles of students at the entry of Kindergarten? (2) Does the membership of the identified HLE profiles vary upon the racial groups and SES? and (3) How does the identified HLE profiles moderate the association between the reading achievement and demographic factors (race and SES)? Study 2 also posed three research questions: (1) What are the SEC profiles of the participants at the kindergarten entry and grade 1 based on the parent-reported and teacher-reported scales? (2) What are the mover-stayer patterns of the identified SEC profiles between the two time points? and (3) What are the long-term effects of SEC on the later reading achievement (grade1) after controlling for SES, gender, and early reading achievement? Finally, study3 aimed to examine the interaction effects between the ecological, cognitive, and psychological factors on identifying the at-risk readers.

Conclusions

This study is a secondary data analysis based on the Early Childhood Longitudinal Study Kindergarten (ECLS-K) 2010-2011.The participants in this study consisted of 13,367 early graders who are in the kindergarten for the first time in 2010-2011 school year and advanced to the first grade in the next year. The reading achievement variables were the IRT-scaled reading scores at the kindergarten fall and first grade spring. HLE measures include 13 survey items regarding direct and indirect literacy interaction between children and parents. SEC measures are the social skill items based on parent-reported and teacher-reported scales for both grade levels. These measures include five items: three positive behavior measures and two negative behavior measures. Three demographic factors measures are socio-economic status (SES), race/ethnicity variable, and gender.

Latent Variable Model and Decision Tree were used for data analysis. Specifically, in Study1, the Latent Class Analysis (LCA) was conducted to address the first research question. The Latent Regression Analysis and the model with distal outcomes were invoked for the second and third research question. Latent Profile Analysis was conducted to identify the SEC profile was conducted in Study 2. The Latent Transition Analysis and the model with both covariates and distal outcomes were used to test the stability of SEC profiles over times and its long-term effect. In Study 3, the Decision Tree Model was employed to identify the specific features of at-risk readers.

The results of this study and the research questions addressed present some validation evidence and important conclusions. In the first study, three home literacy environment profiles were identified: Class1 (pro-reading 1), Class2 (pro-reading 2), and Class 3 (contra-reading). The students classified into Class 1 have the richest home literacy environment: highest likelihood of frequent literacy interaction. Students in Class 2 also indicate their exposure to rich home literacy environment, with exception to the frequency of talking about nature with children. However, the Class 2 students still show higher probability of frequent literacy activities with parents. In contrast, the students belonging to Class 3 exhibit much less likelihood of frequent literacy activities with parents. Hence, based on the Latent Class Analysis, two pro-reading classes and one contra-reading class were identified. Among these three HLE profiles, Class 2 (pro-reading 2) is the most prevalent class (51%), and Class 1 is the next one (33.9%).

The combined prevalence of these two pro-reading groups are almost 85% of the total participants, indicating that most of the sampled family are supportive of literacy activities and have highly frequent literacy interactions between children and parents. And among the pro-reading families, almost 61% of them are less likely to talk about nature at home.

Moreover, overall, the sampled families across the racial groups have the high probability of 70% in supporting literacy activities at home. However, the racial differences were still found as Caucasian families are more likely to provide the supportive and interactive home literacy environment in comparison with other racial groups.

Next, the HLE profiles were also found to vary upon the SES across and within each racial group. High SES families have higher chance of actively getting involved in the literacy activities across racial groups: As the SES grows, the probability of being in the pro-reading HLE increases rapidly, and the probability of being in the contra-reading group drops dramatically. The racial group-specific association between HLE profiles and SES was also found. Compared with other racial groups, the SES, for Asian families, seems to be the more influential factor in predicting the likelihood of the parent's participation to home literacy activities. In other words, the low SES Asian families tend to be less likely to participate in home literacy activities in comparison with other racial groups. This counter-intuitive finding might be due to the items in ECLS-K that are not culturally adaptive. For example, Asian parents generally put more effort and spent more time in providing the in-direct learning environment. They assigned time blocks and gave space for children to study, as well as extracurricular activities during the weekend. While Caucasian parents tended to focus on the in-person supervision: reading books to kids, playing games with kids, and other direct interactions. In short, Asian American parents are more likely to provide indirect home supportive environment than their counterparts. The HLE items available for this study only included those in-person and hands-on activities items. The HLE profiles would be different and comprehensive, if more in-direct support items could be added in the future study.

Finally, the first study also examined the influence of early HLE, racial groups, and SES on reading development. Overall, students in both pro-reading groups (Class 1 and Class 2) outperformed their counterparts in the contra-reading group (Class 3) at both kindergarten entry and the first grade. Across HLE, Asian students had better performance reading achievement in both grades compared with their Caucasian, Hispanic and African-American counterparts. After controlling for SES, the racial gaps in both grade reading achievement between Caucasian and non-Asian counterparts were not noticeable in the contra-reading group. Compared with the demographic factors: SES and race/ethnicity, the initial reading achievement at the kindergarten entry has the largest explanatory power in accounting for the variance in the reading achievement at later reading achievement across HLE profiles.

Study 2 examined the SEC profiles, over-time stability and long-term effect on the later reading achievement. Two types of SEC were identified based on the teacher-reported and parent-reported scales on both grade levels: negative (Class 1) and positive

(Class 2) socio-emotional behavior. Students in the positive socio-emotional behavior group tend to be better at self-control, interpersonal skills, and social interaction, and more likely to exhibit positive approaches to learning, but less likely to have impulsive/overactive behaviors, the feeling of sad and loneliness, and externalizing/internalizing behavior problems. In contrast, students in the negative socio-emotion behavior group have the opposite features: more likely to have problems in self-control, interpersonal communication, and social interaction. Moreover, they tend to feel sad and lonely and exhibit the negative approaches to learning.

The LTA indicated that most of the children stayed in these two SEC profiles. Only 14% of children moved from the negative to the positive socio-emotional behaviors group, while 24% of children exhibited socio-emotional behaviors change from the positive to the negative category, indicating a relatively SEC stability over time.

Furthermore, there was no statistically significant gender gap in the probabilities of being classified as the stayers in both SEC profiles. Thus, the likelihood of being stayers in either SEC profile did not depend upon the gender; so did the likelihood of being classified as the movers from the positive to the negative SEC groups. Conversely, there was a statistically significant gender gap in the probabilities of being moved from the negative to the positive SEC groups. Specifically, female student are more likely to have a socio-emotional skills improvement from negative to positive between the kindergarten entry and grade 1 in comparison with their male counterparts. Moreover, students from higher SES family are more likely to stay in the positive SEC behavior group and transition from negative to positive SEC group, but less likely to stay in the negative SEC behavior group as well as move from positive to negative SEC group. However, this stayer-mover pattern is not gender-specific.

Finally, there was also a significant gap in later reading achievement between these two SEC groups. Specifically, students in the positive SEC group significantly outperformed their counterparts in the negative SEC group after controlling for early reading achievement, gender and SES. The analyses of the second study were conducted on both parent-reported and teacher-reported scales. The results were consistent across these two sets of scales.

The final study investigated the important variables in identifying the at-risk readers based on the Decision Tree model. The final tree indicated that the early reading achievement (kindergarten entry Fall 2010) is the most important predictor in identifying the at-risk readers. Other relatively important variables include teacher-reported SEC status, SES, gender, and parent-reported SEC status. The identified HLE profile, however, was not included in the splitting rules. This might be due to its high correlation with other variables, and thus was possibly masked. The splitting value also profiled the features of at-risk readers from the sampled participants. The highest probability of being classified as at-risk reader is 62.39%. (n=117). The profiles of at-risk students are the early reading achievement lower than 28.104, stayer in or mover to negative SEC behavior group, the SES lower than -.745, Hispanic, male, and stayer or initial status in negative SEC behaviors group. To sum up, the students who are male, from lower SES Hispanic family, have low entry reading achievement had high probability of being identified as at-risk readers. This is because the HLE profiles also significantly differentiated reading achievement at the kindergarten entry. The students with low achievement also refers to the contra-reading HLE profiles. Hence, in addition to the features of at-risk readers, the contra-reading HLE is another feature.

In conclusion, this study examined the effects of the ecological factors (i.e., home literacy environment), psychological factors (e.g., socio-emotional competence) on the early reading achievement. Over 70% of sampled families provided supportive reading environment. Children from the pro-reading families outperformed their counter-reading families. When examined by racial groups, students from Asian families have less home literacy activities with parents than their counterparts of other race, and the SES differentiate the engagement of parents into home literacy activities. However, Asian American children still outperformed their counterparts in reading achievement. Next, students from higher SES families tend to be more likely to experience rich home literacy, and thus obtain early advantage in reading achievement. Moreover, these children are more likely to stay in and move to the positive SEC behavior state in comparison with their counterparts from low SES families, and the negative SEC behavior state was also associated with low reading achievement.

Recommendations

Further research is needed to expand upon with inclusion of survey items from more collection waves as the higher grade levels ECLS-K data will be released in the near future. Furthermore, the study could be improved in terms of research methodologies by invoking complex survey design weights in latent variable modeling.

However, it would highly depend upon the advance of research methods. When the modeling methods are well developed, those methods will be used to validate the results from the current study. Moreover, the decision tree model in the currently could be also further validated by conducting cross-validation study in the near future. Again, the potential issues related to decision tree model with complex survey sampling design data are still unresolved. Further methodological study could be conducted on this topic. Finally, in the third study, the cut-off value for at-risk reader was set at the bottom 10 percent. It is still arguable upon the validity of this fixed value. Hence, further study could be expanded upon by including multiple cut-off values to examine the misclassification rates and explore the impact of selecting different cut-off values on identifying at-risker readers.

REFERENCES

Aaron, P.G., Joshi, R.M., & Quatroche, D. (2008). *Becoming a professional reading teacher*. Baltimore, MD: Paul H. Brookes Publishing Co.

Asparouhov T., Muthén B. (2014). Auxiliary variables in mixture modeling: Three-step approaches using Mplus*.* *Structural Equation Modeling,* 21, 329-341

Blair, C. (2002). School readiness: Integrating cognition and emotion in a neurobiological conceptualization of children's functioning at school entry. *American Psychologist, 57*, 111-127.

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development*, *78*, 647-663.

Bodovski, K., & Farkas, G. (2007). Do instructional practices contribute to inequality in achievement? The case of mathematics instruction in kindergarten. *Journal of Early Childhood Research*, *5*, 301-322.

Bolck, A, Croon, M. A. , & Hagenaars, J. A. (2004). Estimating latent structure models with categorical variables: One-step versus three-step estimators. *Political Analysis.12,* 3–27.

Borghans, L., Duckworth, A. L., Heckman, J. J., & Ter Weel, B. (2008). The economics and psychology of personality traits. *Journal of Human Resources*, *43*, 972-1059.

Bradley, R. H., Corwyn, R. F., McAdoo, H. P., & García Coll, C. (2001). The home environments of children in the United States part I: Variations by age, ethnicity, and poverty status. *Child Development*, *72*, 1844-1867.

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*. New York, NY: CRC press.

Bronfenbrenner, U. (1986). Ecology of the family as a context for human development: Research perspectives. *Developmental Psychology, 22*, 723-742.

Bronfenbrenner, U. (1994). Ecological models of human development. *Readings on the Development of Children*, *2*, 37–43.

Brooks-Gunn, J., & Markman, L. B. (2005). The contribution of parenting to ethnic and racial gaps in school readiness. *The Future of Children, 15*, 139-168.

Buchmann, C., & DiPrete, T. A. (2006). The growing female advantage in college completion: The role of family background and academic achievement. *American Sociological Review*, *71*(4), 515-541.

Bus, A. G., van Ijzendoorn, M. H., & Pellegrini, A. D. (1995). Joint book reading makes for success in learning to read: A meta-analysis on intergenerational transmission of literacy. *Review of Educational Research*, *65*, 1-21.

Chiu, M. M., McBride-Chang, C., & Lin, D. (2012). Ecological, psychological, and cognitive components of reading difficulties testing the component model of reading in fourth graders across 38 countries. *Journal of Learning Disabilities*, *45*(5), 391-405.

Clark, S. L, & Muthén, B. O. (2009). *Relating latent class analysis results to variables* *not included in the analysis.* Retrieved from http://statmodel2.com/download/relatinglca.pdf

Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences*. New York: NY, Wiley.

Compton, D. L., Fuchs, D., Fuchs, L. S., & Bryant, J. D. (2006). Selecting at-risk readers in first grade for early intervention: A two-year longitudinal study of decision rules and procedures. *Journal of Educational Psychology*, *98*, 394-409.

Connell, C. M., & Prinz, R. J. (2002). The impact of childcare and parent–child interactions on school readiness and social skills development for low-income African American children. *Journal of School Psychology*, *40*(2), 177-193.

Gordon, L. (2013). Using classification and regression trees (cart) in SAS® Enterprise Miner for applications in public health. *University of Kentucky, Lexington, KY*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.381.2&rep=rep1&type=pdf

Griffin, E. A., & Morrison, F. J. (1997). The unique contribution of home literacy environment to differences in early literacy skills 1. *Early Child Development and Care*, *127*(1), 233-243.

Dearing, E., McCartney, K., & Taylor, B. A. (2006). Within-child associations between family income and externalizing and internalizing problems. *Developmental Psychology*, *42*(2), 237-252.

Diamond, K. E., Reagan, A. J., & Bandyk, J. E. (2000). Parents' conceptions of kindergarten readiness: Relationships with race, ethnicity, and development. *Journal of Educational Research*, *94*(2), 93-100.

DiPerna, J. C., Lei, P.-W., & Reid, E. E. (2007). Kindergarten predictors of mathematical growth in the primary grades: An investigation using the early childhood longitudinal study--kindergarten cohort. *Journal of Educational Psychology*, *99*, 369–379.

DiPrete, T. A., & Buchmann, C. (2006). Gender-specific trends in the value of education and the emerging gender gap in college completion. *Demography*, *43*, 1-24.

DiPrete, T. A., & Jennings, J. L. (2012). Social and behavioral skills and the gender gap in early educational achievement. *Social Science Research*, *41*, 1-15.

Downey, D. B., & Vogt Yuan, A. S. (2005). Sex differences in school performance during high school: Puzzling patterns and possible explanations. *The Sociological Quarterly*, *46*(2), 299-321.

Drummond, K. V., & Stipek, D. (2004). Low-income parents' beliefs about their role in children's academic learning. *The Elementary School Journal, 104*(3),197-213.

Farkas, G., Grobe, R. P., Sheehan, D., & Shuan, Y. (1990). Cultural resources and s school success: Gender, ethnicity, and poverty groups within an urban school district. *American Sociological Review, 55*(1), 127-142.

Feingold, A., Tiberio, S. S., & Capaldi, D. M. (2014). New approaches for examining associations with latent categorical variables: Applications to substance abuse and aggression. *Psychology of Addictive Behaviors*, *28*(1), 257.

Garcia, E. (2015). *Inequalities at the starting gate: Cognitive and noncognitive skills gaps between 2010-2011 kindergarten classmates*. Washington, DC: Economic Policy Institute. Retrieved from http://www.epi.org/files/pdf/85032c.pdf

Garcia, E. & Weiss, E. (2015). *Early education gaps by social class and race start U.S. children out on unequal footing.* Washington, DC: Economic Policy Institute. Retrieved from http://www.epi.org/files/2015/Inequality-Starting-Gate-Summary-of-Findings.pdf.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, *7*(1), 6–10.

Hamilton, L. G., Hayiou-Thomas, M. E., Hulme, C., & Snowling, M. J. (2016). The home literacy environment as a predictor of the early literacy development of children at family-risk of dyslexia. *Scientific Studies of Reading*, *20*(5), 401–419.

Hecht, S. A., Burgess, S. R., Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2000). Explaining social class differences in growth of reading skills from beginning kindergarten through fourth-grade: The role of phonological awareness, rate of access, and print knowledge. *Reading and Writing: An Interdisciplinary Journal*, *12*(1-2), 99-128.

Heckman, J. J., & Rubinstein, Y. (2001). The importance of noncognitive skills: Lessons from the GED testing program. *The American Economic Review*, *91*(2), 145-149.

Hemmerechts, K., Agirdag, O., & Kavadias, D. (2016). The relationship between parental literacy involvement, socio-economic status and reading literacy.

Educational Review. Advanced online pulication. doi: 10.1080/00131911.2016.1164667.

Herman, K. C., Ostrander, R., Walkup, J. T., Silva, S. G., & March, J. S. (2007). Empirically derived subtypes of adolescent depression: Latent profile analysis of co-occurring symptoms in the Treatment for Adolescents with Depression Study (TADS). *Journal of Consulting and Clinical Psychology*, *75*(5), 716.

Hernandez, M., Folsom, J. S., Al Otaiba, S., Greulich, L., Thomas-Tate, S., & Connor, C. M. (2012). The componential model of reading: Predicting first grade reading performance of culturally diverse students from ecological, psychological, and cognitive factors assessed at kindergarten entry. *Journal of Learning Disabilities, 45*(5), 406-417.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal*, *2*(2), 127–160.

Howse, R. B., Lange, G., Farran, D. C., & Boyles, C. D. (2003). Motivation and self-regulation as predictors of achievement in economically disadvantaged young children. *The Journal of Experimental Education*, *71*(2), 151-174.

Jones, S. M., Brown, J. L., & Lawrence Aber, J. (2011). Two-year impacts of a universal school-based social-emotional and literacy intervention: An experiment in translational developmental research. *Child Development*, *82*(2), 533–554.

Joshi, R. M., & Aaron, P. G. (2000). The component model of reading: Simple view of reading made a little more complex. *Reading Psychology*, *21*(2), 85–97.

Joshi, R. M., & Aaron, P. G. (2012). Componential model of reading (CMR) validation studies. *Journal of Learning Disabilities*, *45*(5), 387–390.

Joshi, R. M., Tao, S., Aaron, P. G., & Quiroz, B. (2012). Cognitive component of componential model of reading applied to different orthographies. *Journal of Learning Disabilities*, *45*(5), 480–486.

Joshi, R. M., Ji, X. R., Breznitz, Z., Amiel, M., & Yulia, A. (2015). Validation of the simple view of reading in hebrew—A semitic language. *Scientific Studies of Reading*, *19*(3), 243–252.

Kieffer, M. J., Vukovic, R. K., & Berry, D. (2013). Roles of attention shifting and inhibitory control in fourth-grade reading comprehension. *Reading Research Quarterly*, *48*(4), 333-348.

Koon, S., Petscher, Y., & Foorman, B. R. (2014). Using evidence-based decision trees instead of formulas to identify at-risk readers. REL 2014-036. Retrieved from http://eric.ed.gov/?id=ED545225.

Kuhn, L., Page, K., Ward, J., & Worrall-Carter, L. (2014). The process and utility of classification and regression tree methodology in nursing research. *Journal of Advanced Nursing*, *70*(6), 1276-1286.

Ladd, G. W., Kochenderfer, B. J., & Coleman, C. C. (1996). Friendship quality as a predictor of young children's early school adjustment. *Child Development*, *67*(3), 1103-1118.

Landa, R. J., Gross, A. L., Stuart, E. A., & Bauman, M. (2012). Latent class analysis of early developmental trajectory in baby siblings of children with autism. *Journal of Child Psychology and Psychiatry*, *53*(9), 986-996.

Laursen, B. P., & Hoff, E. (2006). Person-centered and variable-centered approaches to longitudinal data. *Merrill-Palmer Quarterly*, *52*(3), 377–389.

Lanza, S. T., Collins, L. M., Lemmon, D. R., & Schafer, J. L. (2007). PROC LCA: A SAS procedure for latent class analysis. *Structural Equation Modeling*, *14*(4), 671-694.

Lanza, S. T., Patrick, M. E., & Maggs, J. L. (2010). Latent transition analysis: benefits of a latent variable approach to modeling transitions in substance use. *Journal of Drug Issues*, *40*(1), 93-120.

Lee, V. E., & Burkam, D. T (2002) Inequality at the starting gate: Social background differences in achievement as children begin school. Washington, DC: Economic Policy Institute. Retrieved from http://epsl.asu.edu/epru/articles/EPRU-0603-138-OWI.pdf

Leseman, P. P. M., & De Jong, P. F. (1998). Home literacy: Opportunity, instruction, cooperation and social-emotional quality predicting early reading achievement. *Reading Research Quarterly*, *33*(3), 294–318.

Li-Grining, C., Votruba-Drzal, E., Maldonado-Carreno, C., & Haas, K. (2010). Children's early approaches to learning and academic trajectories through fifth grade. *Developmental Psychology*, *46*(5), 1062–1077.

Lonigan, C. J., Burgess, S. R., & Anthony, J. L. (2000). Development of emergent literacy and early reading skills in preschool children: Evidence from a latent-variable longitudinal study. *Developmental Psychology*, *36*(5), 596–613.

Magnusson, D. (2003). The person approach: Concepts, measurement models, and research strategy. *New Directions for Child and Adolescent Development, 2003*(101), 3–23.

Masyn, K. (2013). Latent class analysis and finite mixture modeling. In T. D. Little (Ed.), *The Oxford handbook of quantitative methods in psychology* (Vol. 2, pp. 551-611). New York, NY: Oxford University Press.

McClelland, M. M., Morrison, F. J., & Holmes, D. L. (2000). Children at risk for early academic problems: the role of learning-related social skills. *Early Childhood Research Quarterly*, *15*(3), 307–329.

McBride, C. (2015). *Children's literacy development: a cross-cultural perspective on learning to read and write*. New York, NY: Routledge.

McClelland, M. M., Acock, A. C., & Morrison, F. J. (2006). The impact of kindergarten learning-related skills on academic trajectories at the end of elementary school. *Early Childhood Research Quarterly*, *21*(4), 471–490

Morgan, J. (2014). *Classification and regression tree analysis*. (Technical report No. 1). Boston University School of Public Health. Retrieve from https://www.bu.edu/sph/files/2014/05/MorganCART.pdf

Morgan, P. L., Farkas, G., & Wu, Q. (2009). Five-year growth trajectories of kindergarten children with learning difficulties in mathematics. *Journal of Learning Disabilities, 42*(4), 306-21

Muthén, B. O. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed), *Handbook of quantitative methodology for the social sciences* (pp. 345–368). Newbury Park, CA: Sage

Muthén, B. O., & Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research*, *24*(6), 882–891.

Muthén, L. K., & Muthén, B. O. (1998-2012). Mplus User's Guide. 7th Edition. Los Angeles, CA: Muthén & Muthén

Nylund, K. L. (2007). *Latent transition analysis: Modeling extensions and an application to peer victimization* (Doctoral dissertation), University of California Los Angeles. Retrieved from https://www.statmodel.com/download/nylunddis.pdf

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling*, *14*(4), 535-569.

Payne, A. C., Whitehurst, G. J., & Angell, A. L. (1994). The role of home literacy environment in the development of language ability in preschool children from low-income families. *Early Childhood Research Quarterly*, *9*(3-4), 427-440.

Phillips, B. M., & Lonigan, C. J. (2009). Variations in the home literacy environment of preschool children: A cluster analytic approach. *Scientific Studies of Reading*, *13*(2), 146–174.

Purcell-Gates, V. (1996). Stories, coupons, and the “tv guide:” relationships between home literacy experiences and emergent literacy knowledge. *Reading Research Quarterly*, *31*(4), 406–428.

Raz, I. S., & Bryant, P. (1990). Social background, phonological awareness and children’s reading. *British Journal of Developmental Psychology*, *8*(3), 209–225.

Reardon, S. F. (2011). *The widening academic achievement gap between the rich and the poor: New evidence and possible explanations*. Retrieved from http://cepatest.stanford.edu/sites/default/files/reardon%20whither%20opportunity%20-%20chapter%205.pdf

SAS Institute Inc (2015). Getting Started with SAS® Enterprise Miner™ 14.1. Cary, NC: SAS Institute Inc.

Scarborough, H. S., & Dobrich, W. (1994). On the efficacy of reading to preschoolers. *Developmental Review*, *14*(3), 245-302.

Scarborough, H. S., Dobrich, W., & Hager, M. (1991). Preschool literacy experience and later reading achievement. *Journal of Learning Disabilities*, *24*(8), 508-511.

Stipek, D., Newton, S., & Chudgar, A. (2010). Learning-related behaviors and literacy achievement in elementary school-aged children. *Early Childhood Research Quarterly*, *25*(3), 385–395.

Tichnor-Wagner, A., Garwood, J. D., Bratsch-Hines, M., & Vernon-Feagans, L. (2016). Home literacy environments and foundational literacy skills for struggling and nonstruggling readers in rural early elementary schools. *Learning Disabilities Research & Practice*, *31*(1), 6-21.

Tourangeau, K., Nord, C., Lê, T., Sorongon, A. G., Hagedorn, M. C., Daly, P., & Najarian, M. (2013). *Early Childhood Longitudinal Study, Kindergarten Class of 2010-11 (ECLS-K: 2011), User's Manual for the ECLS-K: 2011 Kindergarten Data File and Electronic Codebook* (2013-061). U.S. Department of Education. Washington, DC: National Center for Education Statistics

Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. In J. A. Hagenaars & A. L. McCutcheon (Eds.), *Applied latent class models* (pp. 89-106). Cambridge, UK: Cambridge University Press.

Vermunt, J. K., & Magidson, J. (2005). Latent Gold 4.0 User's Guide. Belmont, MA: Statistical Innovations Inc.

Vermunt J.K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis.18,* 450–469.

Waldfogel, J. (2012). The role of out-of-school factors in the literacy problem. *The Future of Children*, *22*(2), 39–54.

Wheaton, A. (2010). *The home literacy environment of children in kindergarten through first grade as a predictor of third grade reading abilities* (Unpublished Master Thesis). University of Arkansas.

Yarosz, D. J., & Barnett, W. S. (2001). Who reads to young children?:Identifying predictors of family reading activities. *Reading Psychology*, *22*(1), 67-81.