**Title**
Predicting Students' English Performance with Traditional Statistical Modeling and Machine Learning: An Analysis of the China Education Panel Survey (CEPS)



**Author**
Xia, Yuzhu

**Publication Date**
2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
SANTA CRUZ

**PREDICTING STUDENTS' ENGLISH PERFORMANCE WITH TRADITIONAL STATISTICAL MODELING AND MACHINE LEARNING: AN ANALYSIS OF THE CHINA EDUCATION PANEL SURVEY (CEPS)**

A dissertation submitted in partial satisfaction
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

EDUCATION

by

**Yuzhu Xia**

June 2022

The Dissertation of Yuzhu Xia is approved:

Professor Kip Téllez, chair

Professor Eduardo Mosqueda

Professor Douglas Bonett

Peter Biehl
Vice Provost and Dean of Graduate Studies

# Table of Contents

## List of Figures and/or Table of Illustrations

## Abstract

Predicting students' English performance with traditional statistical modeling and machine learning: An analysis of the China Education Panel Survey (CEPS)

Yuzhu Xia

With the global expansion of English teaching, factors related to language achievement have recently garnered a significant amount of attention (Onwuegbuzie, et al., 2000; Phillipson & Phillipson, 2007). This research aims to contribute to the literature on English achievement in the Chinese context by examining the influence of specific key variables (e.g., students' grade level, parent involvement, teacher characteristics, school demographics) on English achievement scores. The data are taken from the China Education Panel Survey (CEPS), a large-scale, nationally representative, longitudinal survey starting with two cohorts (7th and 9th graders enrolled in the 2013-2014 academic year). In addition to exploring English achievement, the study also contributes to the literature on quantitative methodologies in the context of educational research by exploring the use of statistical modeling and machine learning in studies on academic achievement. Analyses from both multilevel modeling and Support Vector Regression (SVR) revealed that students' English performance was largely explained by their scores on Chinese language performance, cognitive aptitude scores, self-perceived educational expectations, and parents' expectations of their children's academic

performance and future educational achievement. The current study corroborates the findings of previous research, which demonstrate that achievement in one’s native language is associated with the achievement of languages learned later in life (Ortega, 2014). Both multilevel modeling and SVR were shown to be useful methods for predicting English achievement, suggesting that educators and researchers may benefit from both approaches to further understand the broader variable of academic achievement.

## Acknowledgments

I would like to thank my advisor Kip Téllez, who offered me a great amount of guidance, understanding, and support in my academic and personal life. A huge thanks goes to my other dissertation committee members Eduardo Mosqueda and Douglas Bonett, whose passion and wisdom in statistics and social science inspired me to pursue the quantitative social science track in education. I thank them for their help with my dissertation and for the support and understanding they have provided me with throughout my graduate journey. I would also like to thank my lovely friends and colleagues at UC Santa Cruz – Liwei Liu, a lifelong friend with whom I spent most of my time on campus, and the MALLI crew – Jolene Gregory, Adria Patthoff, and Liana Balloffet. In addition to having enjoyed working as a team with them on research projects and academic papers, I loved to simply hang out with them as well. I am especially grateful to have the most supportive and loving family: my mom, who has always been my role model for being a nice and warm person; my husband Zifeng, who took over most of the family responsibilities when I was occupied with work and grad school, and our son Ray, who gifts me an abundance of smiles and warm hugs exactly when I need them. I would not have been able to complete this dissertation or the graduate degree without the academic and emotional support from my graduate program, research team, all my friends, colleagues, and family. I feel blessed to have you all in my life.

## Introduction

China has the largest education system in the world, serving over 260 million students and employing more than 15 million teachers (OECD, 2016). According to the 2019 National Bureau of Statistics of China, there are 161,811 primary schools, 51,982 middle schools, 13,737 high schools, 10,299 secondary vocational schools, 2,663 higher education institutions, and 1,418 postsecondary vocational institutions in China. Educational research in China has experienced a complex and circuitous journey ever since. Chinese educators and scholars first began a "westernization" of educational theories and methods at the start of the 20th century. It is only in recent decades that this field of research began adopting recognized scientific research methods (Wen & Xie, 2017). Educators in China have also realized their need for indigenized educational theories (Ye, 2004) as potential concerns with the use of westernized theories to address local problems have started to emerge (Wen & Xie, 2017; Yang, 2005).

As large-scale educational datasets are presently available in China, educators and researchers consider the need for further research, especially investigations that employ quantitative methods to analyze and interpret data, as this may potentially lead to the formation of novel educational theories (Yue & Xu, 2019; Zhao, et al., 2008). A very limited number of studies have adopted advanced modeling methods to examine educational datasets (Yue & Xu, 2019). The present study examined the China Education Panel Survey (CEPS) dataset that was publicized in recent years and contributed to the field of English achievement in China by offering significant predictors. Furthermore, the study makes a notable contribution to the field of quantitative educational research by offering new perspectives and exploring the roles of traditional statistical modeling and machine learning in generating models and predicting student performance in English.

## Literature Review

### English Language Education in China

English language education has been regarded as both a personal and national asset in China since the last quarter of the 20th century, given the broader political and economic context of modernization and development (Hu, 2005). English was mandated as a compulsory subject in all secondary schools from the beginning of the late 1970s; later on expanding to the primary sector at the start of the 20th century, following the call of the Ministry of Education (MOE) (Hu, 2002a). In order to pursue political and economic advancement after the opening of new China in the late 1970s, English was used as a medium, on a national level, to educate more high-quality human resources that meet the global standard to provide efficient contributions to the modernization of Chinese society. At the individual level, English was deemed an essential skill for success in academic life and careers (Hu, 2005, Wang &

Gao, 2008). Moreover, education has broadly been regarded by several as “a panacea for social and personal problems” (Thøgersen, 2002, Wang & Gao, 2008). Compounded by these contextual observations, English achievement became one of the few important standards that families, schools, and the Chinese society at large adopted to determine the achievement and quality of the workforce.

Extant literature examining English language education in China generally covers six major areas: 1) the historical discussion and overview of English language education; 2) the usage of the English language in China; 3) language policy and planning; 4) curriculum reform initiatives and implementation; 5) learner experience; and 6) the teaching workforce, with some areas corresponding with a very limited amount of research (Wang & Gao, 2008).

Overall, the rise of China as an empire was greatly facilitated by an improvement in the overall education quality over time and the enforcement of English as a compulsory subject in the Chinese educational system. China has been through nine English curriculum reforms in recent decades; notably, the allotted class hours for English are significantly higher than the mandated number initially drafted in the national curriculum (Hu, 2005). Relatedly, a dramatic expansion of the English teaching workforce occurred alongside the observance of conundrums such as low quality of teaching and lack of professional background (SEC Department of Planning and Construction, 1991). Due to the global developments in terms of the propagation of English as a studied subject along with the various issues associated with the developments, educational researchers in China have endeavored to study English education overall and issues that have emerged, specifically, English achievement and potential indicators of students' English achievement as one line of research. In a broader sense, foreign language achievement is associated with various cognitive, affective, personality, and demographic predictors (Onwuegbuzie, et al., 2000), in addition to academic expectations and beliefs (Phillipson & Phillipson, 2007). Limited empirical studies within the Chinese context also show similar results (Li, et al., 2012; Wang, 2008; Wen & Johnson, 1997). However, several gaps in the literature have been identified that require to be addressed, such as: further consideration of the linguistic context in China, a closer examination of the dynamic learning and teaching practices at the individual and the class level in both public and private settings. This study intends to contribute to the body of literature on English achievement in China by identifying some accurate predictors of students' English scores in middle schools as well as the effective data analysis approaches they require.

## Second Language Acquisition (SLA)

Part of the analysis in this study is grounded in the theory of second language acquisition (SLA). With the development of SLA over the past few decades, scholars have been prompted to focus on "the nature of the language acquisition process and the factors which affect language learners" (Larsen-

Freeman, 1991). By studying the former, cognitive approaches and sociocultural approaches emphasize various routes and factors that influence and/or determine the SLA process (Zuengler & Miller, 2006). Cognitive approaches to SLA emphasize universal grammar (Chomsky, 1965, 1981, 1986). As Ellis (1999) concluded, these approaches regard linguistic signs “as a set of mappings between phonological forms and conceptual meanings or communicative intentions” (p. 5). Ultimately, these approaches have contributed to a large body of research on several sub-areas including Functional linguistics (Bates & MacWhinney, 1981), Emergentism (Elman, Bates, & Johnson, 1996), Cognitive linguistics (Langacker, 1987, 1991), and Constructivism (Slobin, 1997). The cognitive approaches generally agree on the notion that language acquisition is determined by the brain processing of the language as well as individual short-term and long-term memory (Ellis, 1999, 2008).

A sociocultural perspective, however, rejects the notion that the language acquisition process is purely psychological or cognitive (Bakhtin, 1981; Lave & Wenger, 1991; Ochs, 1988; Vygotsky, 1978; Watson-Gegeo, 2004). Sociolinguists argue that social factors and situational factors can affect the language acquisition process of learners, which include but are not limited to: age, gender, social class, and ethnic identity. This study adopted the sociocultural perspective to examine students’ English performance by first considering variables such as the students' age, gender, financial background, and their parents' educational background.

The second focus of SLA researchers, as suggested by Larsen-Freeman (1991), is placed on the factors affecting learners' SLA process. Extant literature suggests that comprehensible language input is essential to the process of SLA. Krashen's (1982) work on acquisition-learning further explains this form of language input by distinguishing between "acquisition" and "learning", which are two varying processes – the former refers to how children acquire a language in a manner similar to the acquisition of their first language, characteristic of a subconscious process; while learning refers to a conscious process of obtaining knowledge regarding the second language, generally achieved in school settings or the like (Krashen, 1982). In the context of Chinese middle schools, students' exposure to English generally takes place at school; hence, their SLA process is achieved through "learning" in general. The study by Chihara and Oller (1978) proves that formal study is positively correlated to second language proficiency, although there is no consensus within the literature regarding this relationship (Krashen, 1982).

Krashen (1982), in his descriptions of the affective filter hypothesis, argued that a series of affective factors are also related to the SLA process. Proposed by Dulay and Burt (1977), the concept of the affective filter consists of three important categories: motivation, self-confidence, and anxiety (Krashen, 1982). Another proposal was offered by Halliday (1975, 1978, 1993), which suggested that language should be considered as a social semiotic system and that language learning must extend beyond grasping grammar and structures, but also satisfy the various social-functional needs of the learner, which include construing experiences around and within us, negotiating social roles and attitudes to interact with the social world, and creating messages with “themes”, which he defined as ideational, interpersonal, and textual social functions. Students’ foreign language learning and performance in the context of this dataset also involve their beliefs regarding the social functions of the English language, such as their future jobs, locations of work, and desired level of education.

Furthermore, a discussion of learners’ English performance necessitates considering the influence of their first language (L1) (Jarvis & Pavlenko, 2008). Ortega (2014) specifically discussed the cross-linguistic influences in analyzing the acquisition process of any additional language. According to Ortega (2014), L1, or any previous language knowledge that was acquired before one starts to learn an additional language, is an important source of influence on foreign language learning; this “holds universally true of all L2 learners.” The impact of L1 on L2 acquisition has been well-studied in various aspects, such as in terms of phonological inventory and lexical skills (Harrison & Kroll, 2007; Proctor et al., 2006). Researchers have successfully proven that students’ L1 skills can transfer to L2 during L2 acquisition as a subconscious process (Dulay et al., 1982) since learners establish their own rules regarding L2 acquisition with the help of their L1 knowledge and learning skills. This is especially true in the Chinese context since both L1 and L2 are tested in the form of paper-based exams, which, to some degree, magnifies the role of learning habits and rules. Specifically, Proctor et al.'s study (2006) revealed that the stronger the students' L1 vocabulary skills, the better their L2 performance is predicted to be. Empirical studies have also revealed several other important predictors of foreign language achievement other than L1; for example, one's linguistic background (native language), which may be monolingual or bilingual/multilingual (Maluch, et al., 2015), language anxiety (Aida, 1994), and motivation/attitude variables (Dörnyei, 1990).

In the CEPS dataset, various variables from the student, parent, and teacher questionnaires were sourced to measure the above-mentioned concepts. For instance, the students' Chinese score, their attitudes towards their English teacher, if they felt English was difficult for them, and students' feelings regarding their parents' expectations. Furthermore, to test the relationship between these affective factors, data on linguistic background and the students' English performance were collected as well.

### Teacher Influence in Relation to Foreign Language Achievement

Various social and situational factors can affect students' foreign language achievement, including teacher quality and foreign language instruction. An extensive amount of literature has studied the subject of student performance, focused specifically on students' foreign language grades in relation to student-related factors, teacher-related factors, and school-related factors (An, Hannum, & Sargent, 2008; Dossett & Munoz, 2003). A line of research investigating teacher-related factors has focused specifically on teacher quality (An, Hannum, & Sargent, 2008; Buchmann & Hannum, 2001; Darling-Hammond, 2002; Greenberg, et al., 2004; Greenwald, Hedges & Laine, 1996). Research on how teachers and teacher quality can affect student performance dates back to the Equality of Educational Opportunity report, also known as the “Coleman Report” (Coleman et. al., 1966; Huang & Moon, 2009). Since then, educational researchers have documented the associations between teacher characteristics and student performance. Attempts have been made to identify the teacher attributes that influence the effectiveness of their teaching, which subsequently affect student performance. Despite the international significance attributed to student performance, limited research has been identified outside of the U.S. that examines the relationship between teacher attributes and student achievement (Huang & Moon, 2009).

Specifically, however, researchers have examined the relationship between teacher experience and student performance (Ladd & Sorensen, 2017; Rice, 2003; Wayne & Youngs, 2003). Findings of existing literature indicate that a vast amount of experience for a teacher is typically associated with higher teacher quality and a positive effect on student performance and student behavior (Huang & Moon, 2009; Ladd & Sorensen, 2017; Rice, 2003). In defining teacher experience, scholars categorize teachers’ developmental stages into: a beginning stage, one or two developing stages in the middle, and a mature stage when a teacher has acquired five or more years of experience (Burden, 1982; Christensen, et al., 1983; Katz, 1972). For example, Katz (1972) argued that there are at least four developmental stages for teachers: Survival, Consolidation, Renewal, and Maturity. Berliner (1988) used a similar categorization, adding a fifth “expert” stage which only a select few teachers in the maturity stage can reach. Katz’s (1972) categorization was adopted in the context of the present study as her definitions of the four stages present the correspondence between teachers’ professional and psychological development in more detail, and a more specific division of the Consolidation Stage and the Renewal Stage depicted teachers’ professional career path in a relatively more organic manner. In describing the four stages, Katz (1972) admitted that individual teachers may vary in terms of the duration spent at each stage; but generally, the Survival Stage is the first year of teaching, when teachers acquire baseline information about children and need the most “support, understanding, comfort, and guidance” (p. 51); the Consolidation Stage is roughly year 2 and 3, when teachers “begin to focus on individual children who pose problems and on troublesome situations” (p. 51); the Renewal Stage is roughly year 3 and 4, when teachers start to familiarize themselves with the classroom and teaching and require renewal and refreshment by exchanging experience with colleagues or reflecting upon their own teaching strategies; the Maturity Stage is usually achieved by teachers after five years of teaching, when they start to develop a more comprehensive reflection of themselves as teachers and begin posing deeper questions regarding their teaching and teaching as a career.

The differences in attributes and capabilities of teachers at different developmental stages are directly linked to their practices in classrooms. Teachers at the Survival Stage are typically concerned with completing individual tasks and getting through daily tasks, which is in line with existing literature that documents the experiences of first-year teachers (Feiman-Nemser, 2012; Fuller, 1969; Fuller, Parsons, & Watkins, 1973; Lortie, 1966). Teachers who reach the Maturity Stage tend to feel more secure and confident. Having realized the complexities of children's academic and developmental needs, mature teachers gradually become able to "adopt a more child-centered approach" (Burden, 1982). The stages between Survival and Maturity are understood as the adjustment stage or the middle years. During this time, teachers are in the process of developing their teacher sense in the classrooms, honing their teaching skills to perfection, and attempting to attend to children's needs more freely. Therefore, teachers at different developmental stages can have varying levels of influence on their students and academic achievement.

Another means by which to describe and investigate teacher experience and the nature of pedagogical expertise is the comparison of targeted aspects of teaching between expert teachers and novice teachers, or between mature teachers and beginning teachers in the perspective of teacher developmental sequence (Xia, 2019). Existing studies that examine the difference in perceptions and instructions between expert and novice teachers shed light on the structure of the expert-novice distinction (Borko & Livingston, 1989; Carter, Sabers, Cushing, Pinnegar, & Berliner, 1987; Chi, Feltovich, & Glaser, 1981; Fogarty, Wang, & Creek, 1983). The current study employs the structure proposed by Borko and Livingston (1989) to analyze the similarities and differences characterized by expert and novice teachers. Borko and Livingston (1989) offer powerful theoretical foundations for examining pedagogical expertise by viewing teaching as a complex cognitive skill and an improvisational performance as well (Shulman, 1987; Yinger, 1987). They argue that teaching is a complex cognitive skill that demands teachers to not only possess content knowledge and pedagogical knowledge but also effectively transform their knowledge into forms that are pedagogically powerful and adaptive to students – this skill requires years of experience to practice (p. 474). Additionally, teachers must also be able to improvise in an ever-changing dynamic environment based on student performance cues, which also requires experience to establish effective classroom routines and agendas (Borko & Livingston, 1989; Leinhardt & Greeno, 1986; Shulman, 1987; Yinger, 1987). Yinger (1987) suggested that teaching, to a great extent, represents an improvisational performance; the performer/teacher must begin with an outline of their performance/teaching (e.g. a rough script or lesson plan for the class) and the detailed steps can only be filled in as the class progresses, based on students' feedback and teacher’s checking of their understanding (Yinger, 1987). Fogarty, Wang, and Creek (1983) also stated that “expertise in semantically rich domains involves the ability to apply knowledge effectively in response to environmental cues” (p. 22).

Another reason for choosing the Borko and Livingston (1989) framework is that they provide clearly structured and comprehensive layers of comparison in planning, interactive teaching, and post-lesson reflections, allowing us to compare and contrast the similarities and differences characterized by expert and novice teachers. Firstly, expert teachers and novice teachers usually plan lessons differently - the former with a simpler outline, trying to utilize student examples generated in class for illustration to make their lessons more relevant to the students; the latter with a more scripted plan and detailed activities to achieve their goals. Westerman (1991) also discovered that expert teachers are masters of integrating knowledge of subject content and prior knowledge to map out new lessons, and are more likely to think from their students’ perspectives. Secondly, due to their difference in their levels of experience, the two groups of teachers show different reactions and levels of flexibility in handling student questions and deviations from their planned structure, as well as in their ability to maintain their instructional goals and pacing in class. Lastly, expert teachers tend to focus more on student learning and performance when they reflect on their lessons. While novice teachers also reflect on student learning, they reflect upon their own teaching as well to develop in terms of personal growth. Borko and Livingston (1989) claimed that the differences “rest on differences in knowledge, which can, in turn, be analyzed in terms of cognitive structures”, as novices’ cognitive schemata are less elaborate, interconnected, and accessible than that of experts (Borko & Livingston, 1989, p. 490; Anderson, 1984; Livingston & Borko, 1990; Shavelson, 1986). Expert teachers are more capable of using information from existing schemata for a new lesson, predicting areas where students are most likely to have problems, and better-informed judgment calls regarding information that is relevant to a lesson and information that may be ignored (Borko & Livingston, 1989). In other words, although expert and novice teachers may carry out similar teaching activities, they may be regarded as different in that they serve their goals at varying levels. As Leinhardt (1986a) put it, “although expert teachers do many of the same things well, they do not necessarily do them in the same way” (p. 33).

Another line of research investigates the influence of foreign language formal instruction on student performance. It is generally agreed that informal and formal environments contribute to varying aspects of second language acquisition. Krashen (1976) further hypothesized that formal study could be significantly more efficient than informal exposure in increasing second language proficiency among adults (p. 158). Various studies have proved that formal foreign language instruction is positively correlated with higher proficiency, while increased informal exposure may not necessarily be linked to higher proficiency (Carroll, 1967; Krashen & Seliger, 1975; Krashen, Seliger, & Hartnett, 1974; Ellis, 1990). In this study, variables such as teachers' preparation, instructional approaches, teaching materials, and teacher attributes (e.g. teacher experience) were included to evaluate their influence on students' English performance.

### Parent Influence on Academic Achievement

Parental involvement and expectations have been adequately addressed with regard to their influence on children's social-emotional and academic standing (Anderson & Minke, 2007; Gutman & Midgley, 2000; Henderson, 1987; Maccoby & Martin, 1983; Phillipson & Phillipson, 2007; Sears, Maccoby, & Levin, 1957). The Coleman Report (Coleman et. al., 1966), in addition to Mosteller and Moynihan's (1972) reanalysis, revealed that about 50% to 67% of the student achievement variance could be explained by home variables (Greenwood & Hickman, 1991). Anderson & Minke (2007) also reported that an increase in parent involvement is associated with improvements in overall academic achievement (Shaver & Walls, 1998), homework completion (Cancio, West, & Young, 2004), and statewide assessment scores (Sheldon, 2003).

The impact of parent expectations, as a specific area of parent involvement in children's academic life, has not been significantly documented in English and Chinese literature. Existing English literature on examining parent involvement has been primarily focused on parenting styles (Spera, 2005; Turner et al., 2009), parent-teacher relationships (Hughes & Kwok, 2007; Minke et al., 2014), types of parental involvement in the academic context (Hill & Tyson, 2009; Jeynes, 2012), and factors such as socioeconomic status and parents' educational background as predictors of greater importance (Caldas & Bankston, 1997; White, 1982).

The limited literature discussing the importance of parent expectations initiated this conversation in the broader context of academia. Parent expectations, as defined by Christenson et al. (1992), are future aspirations or current expectations for children's academic performance. White's (1982) analysis of 101 studies revealed that parent expectations are among the few factors under the umbrella category of parent involvement that can impact student achievement more than SES. Among the existing literature documenting parental expectations in the Western realm, most have focused on elementary school and white students, and the statistical methods used were mostly correlational studies (Christenson et al., 1992). In the Chinese literature, most studies were of a descriptive nature and in formats of survey reports on differing scales. This ties into the specific issues with current educational research in China, as discussed in this dissertation.

## Cognitive Capacity and Language Acquisition/Achievement

Cognitive capacity, otherwise termed aptitude, is an important concept in both the theoretical and practical aspects of English achievement. On the one hand, aptitude carries significant meanings as it measures and revealed factors involved in the process of learning; several researchers argue that cognitive capacity functions as a good indicator of academic achievement, which is measured by standardized test scores (Genesee, 1976) and other important factors (Dockrell & Brousseau, 1967; Gardner & Lambert, 1972). On the other hand, teachers and educators consider it when deciding upon the type of instruction and materials to use, and students are oftentimes measured by this type of test to be accepted into training programs (DeKeyser & Koeth, 2011). The early works of Jim Cummins (Cummins, 1979; 1999) also focused on the association of cognitive aptitude/demand with language development. Therefore, it is of great importance to understand the element of aptitude as well as the viewpoints scholarly literature offers on the same.

Scholars define cognitive capacity, or aptitude, as the characteristics that a learner brings to the learning process, with a general emphasis on the holistic perspective of a learner's capacity or intelligence (Cronbach & Snow, 1977; Carroll & Sapon, 1959; DeKeyser & Koeth, 2011). Language aptitude, as a line of research on the intersection of cognitive capacity and language achievement, started to emerge in the second half of the 20th century (Ameringer, et al., 2018). Carroll (1958, 1964) defined the term as the rate of acquisition in foreign language acquisition. Language aptitude entails a wide range of components that explain an individual's cognitive capability, of which working memory is a widely known concept. As another key factor in language achievement and an important identifier for individual differences (IDs) (Dörnyei, 2006), working memory has been demonstrated to be positively associated with language aptitude and academic achievement (Ameringer, 2018; Baddeley, Gathercole, & Papagno, 1998; Miller, Galanter, & Pribram, 1960; Paivio, 1986). Language aptitude can also help to analyze the potential individual differences and ranges in foreign language learning processes across individuals, such as pacing and progress (Ortega, 2014).

Carroll (1964) proved that aptitude is general across different languages, and the aptitude tests could offer diagnostic possibilities. He also tested the prediction of success using a battery of psychometric tests with relatively high validity coefficients and multiple correlations as high as 0.84. He also provided a model that effectively describes the relationships among aptitude, the ability to understand instruction, motivation, the time allowed for learning, and quality of instruction. Admittedly, cognitive capacity or aptitude testing has always been controversial. Scholars across disciplines have raised the issue of negative connotations these types of tests could bring. Nevertheless, cognitive aptitude is undoubtedly a key factor in explaining language processing, acquisition, and achievement (Jessner, 2006; Singleton, 2017).

## Educational Research in China

Having briefly reviewed the literature on English education in China, SLA, teacher influence, teacher experience, parent influence, and individual characteristics such as cognitive capacity, it is now imperative to streamline the scope of the literature to works in the context of educational research in China

– as this is most relevant to the present study. Educational in research China has been shaped by the unique history, established education system, and scholarly traditions of China, resulting in unique challenges that demand educational theories and practices that are local to its context.

Educational research in China has experienced a complex and circuitous journey in accordance with its drastic social, cultural, political, and economic changes over the past century (Wen & Xie, 2017). While reviewing the complete history of the Chinese educational system and its association with educational research is not the purpose of this study as each period of the context is determined by the dynasty and the needs of the ruling class therein, this contextual background will help us better understand the current status, strengths and weaknesses, and the needs of educational research in China. To this end, I shall briefly summarize the developments in the education system and educational research in China over the past hundred years.

*Development Stages of Educational Research in Modern China*. The first stage spans from the start of the 20th century to the founding of new China (Wen & Xie, 2017). At the turn of the century, China began to send students and scholars overseas, primarily to the US and Europe, to obtain education and return with valuable knowledge of theories and technology in various disciplines, including education, as documented in the report (1954) by Mei Yiqi and Cheng Qibao. At this point in time, education was regarded as a salvation to the country, much more than the disciplines of science and engineering, which would protect them from the potential of invasion (Wen & Xie, 2017; Zhu, 1988). Respected educators and scholars such as Hu Shi and Tao Xingzhi were promoters and pioneers of John Dewey in China. They disseminated empirical study-based and evidence-based research methodologies in the country, preceding which non-empirical and interpretative approaches were employed in the Chinese context of research. Wen and Xie (2017) called this installation of westernized theory in education, accompanied by other social science disciplines, a "shortcut" (p. 148) – these westernized scientific research methodologies reshaped educational research and scholarship.

The second stage was significantly influenced by political ideology, ranging roughly from the founding of new China in 1949 to the end of the cultural revolution in 1976 (Wen & Xie, 2017). This stage marked a shift from learning in the Soviet style to an approach that required one to look inward as a result of the complicated relationship between China and the Soviet Union in the late 1950s and the early 1960s. Despite this influence, the educational research in this period was closely tied with the political ideology of the China Communist Party (Wen & Xie, 2017). In other words, social science disciplines, especially education, simply became tools of the government in the propagation and enforcement of central ideologies. All scientific research in these fields was almost completely suspended.

The third stage spans from after the cultural revolution until the end of the 20th century (Wen & Xie, 2017). This is when education across all levels was rejuvenated, with great developments and innovations taking place in educational research as well. It was not until 1983 that education officially became a discipline, marking the start of a phase of steady development. Accompanying the marketization policy in 1978, educational research gradually adopted performance-based approaches, including empirical studies with quantitative methods; this characterized the dominant research trend in the field of education for the two decades that followed, until the mid/late 1990s, when qualitative methods reclaimed the spotlight (Wen & Xie, 2017).

The last stage ranges from the early 21st century to the present time, marking a period for mass globalization and indigenization of various disciplines (Wen & Xie, 2017). In contemporary educational research, scholars have been debating if China should continue to import westernized theories and practices as it did a century ago, or if it should develop its own theories. This debate is informed by the observation of youths tending to adopt westernized theories and methods from education as tools for application in and interpretation of the Chinese context – which can be a challenging and complicated task.

*Key Features of Educational Research in China*. Most modern educational theories and frameworks applied in the education system and educational research are western-based, introduced in the early 20th century (Wen & Xie, 2017). The lack of indigenized educational theories to guide research is a major issue encountered by educators and scholars in China (Ye, 2004). The epistemological traditions of Western countries and China are significantly different, as the West holds a relatively more analytical view that employs empirical investigation methods, while China adopts the practice of Confucian introspection that values reflection and wisdom (Shun & Wong, 2004; Wong, 2020; Zhao, et. al., 2008). A comparative study investigating about 300 articles revealed that over 90% of published articles in AERJ were concerned with issues guided by theories, while only about a third published in JYYJ (Jiaoyu Yanjiu, or Educational Research, the most prestigious journal in the field of Education in China) were domestic issues; other articles found therein were focused on general educational theories, international issues, and comparative studies, among other subjects (Zhao et. al., 2008). Wen and Xie (2017) concluded that “China does not have its own educational scholarship in a modern sense” (p. 156).

The Chinese government, specifically the Ministry of Education, plays a key role in steering and funding educational research in a top-down fashion (Chen, 2012; Wen, 2005; Wen & Xie, 2017). Reflecting upon the start of the 20th century, the government was observed to have supported the overseas education of students and scholars to return with westernized theories and methods. After the founding of New China, educational research largely followed the government’s policies, becoming a tool for political propaganda. After China reopened in the late 1970s, modern educational research resumed; however, the government continued to determine its development as it controlled the allocation of funding and enrollment quota in research universities, the leading players in scientific research (Wen & Xie, 2017). Even today, the majority of educational research responds to the government's call for educational reforms and other policy-related matters.

Recently, many educational research projects are found to be practice-based rather than driven by theory (Wen & Xie, 2017). According to Zhao et. al. 's (2008) study, of all the articles they studied, 55% of the published articles in JYYJ in China were conceptual papers, nearly 30% of the articles were commentary/self-reflection/historical discussion/policy reports, and only less than 15% of the articles were research projects. In contrast, 93% of the articles published in AERJ were research papers authored by university professors and other scholars.

Relatedly, it has been noted that quantitative research only takes up a small portion of all published articles. Zhao et. al.'s (2008) study revealed that the vast majority of published articles in JYYJ did not adopt widely recognized research methodologies, such as qualitative methods (e.g. ethnographic research), quantitative methods (i.e., experimental/quasi-experimental studies), or mixed-methods; only two of all articles were found to have used experimental/quasi-experimental design, while nine were papers that employed secondary analysis. Other papers were found not to have any original data collection or analysis. This finding coincides with an earlier work of research that examined studies published in the 1980s to 1990s, and found rarity among empirical studies as well (Zheng & Cui, 2001). Yue and Xu (2019) further noted the urgent need for quantitative data mining and analysis with more recently available large datasets.

*Future Directions*. While educational research in China has experienced great development after decades of effort, there remains room for improvement. Firstly, future educational research should be dedicated to developing its own theories and frameworks originating from the Chinese context. These theories must be applied to interpret issues emerging from the local context and propose further steps, which require the identification of gaps and limitations that cannot be fully resolved by western educational theories. Currently, educational research in China continues to adopt westernized theories to resolve local concerns; however, scholars have realized the issues that arise from this approach and are strategizing the means by which they may be resolved.

Secondly, educational research in China must further utilize scientific research methods to conduct their studies. Currently, a large portion of educational studies are self-reflections, commentaries, historical discussions, and policy-related reports. Only a small percentage of published studies have used widely recognized research methodologies such as quantitative, qualitative, or mixed-methods (Yue & Xu, 2019; Zhao et. al., 2008). Wen and Xie (2017) noted that despite the development of quantitative methodologies in the recent decades, qualitative methods started to gain popularity among educational scholars starting from the mid-1990s; however, a dearth of quantitative studies has been noted in the field of education. Studies using educational data mining (EDM) and machine learning are also primarily based in the West (Baker & Yacef, 2009). The few studies that use machine learning in China only started to emerge in the 1990s, but as of yet, a very limited number of quantitative studies have used advanced methods such as machine learning (Yue & Xu, 2019).

With the availability of big datasets for global education purposes, more scholars may be encouraged to use these datasets to conduct quantitative studies (Yue & Xu, 2019). The China Educational Panel Survey (CEPS) is one of the public, large-scale, nationally representative datasets that may enable educators and scholars to develop original education theories – yet this dataset remains understudied. More research articles studying the CEPS dataset have been published in recent years to examine the following topics: 1) student performance; 2) educational inequality; 3) bullying; 4) teaching workforce; 5) gender differences; 6) students' mental well-being; and 7) other areas such as obesity and housework. A large number of studies examine student performance (about over a third), while popular areas of the studies include family background or structural factors associated with student performance (Tani, et al., 2021), math performance, and gender effects (Liu, 2018); classroom composition, e.g. migrant/local students, low-ability/achievement and regular ability/achievement students (Wang, 2021); school hours and structure (Yang & Zhao, 2021); and teacher quality, which encompasses teacher-parent relationship, gender, and education level (Liu, 2021). Other related topics include the urban/rural divide (Sun, 2020), the role of tutoring and extracurricular classes/activities (Guo et al., 2020), achievement gap by gender (Luo et al., 2021), school socioeconomic composition in relation to student achievement (Yuxiao & Chao, 2017), parent involvement (Duan, 2018), and cognitive ability (Li et al., 2019).

Thus far, existing research has failed to examine middle school students' English performance by analyzing the CEPS dataset, determine some good predictors of English achievement scores, and identify methodological approaches that need to be adopted by educational researchers. By analyzing the students' performance and related factors through traditional statistical modeling methods and machine learning, the results of this study contribute to the existing literature by providing good indicators of students' English performance, which will initiate a discussion on the modeling method(s), either traditional statistical modeling or machine learning, or a combination of the both, that may yield better predictions and contribute to the establishment and formation of authentic Chinese educational theories that are unique to the Chinese context.

## Machine Learning and Educational Research

Machine learning has been applied in the sciences since Samuel (1959) defined the term as a field of study wherein computers may learn without the need to be explicitly programmed. Yet, the concept remains understudied across many disciplines, one of which is education. Data mining functions as an information source for machine learning to pull from. Educational data mining (EDM), as defined by Romero and Ventura (2010), is “an emerging interdisciplinary research area that deals with the development of methods to explore data originating in an educational context,” which uses “computational approaches to analyze educational data in order to study educational questions” (p. 1). EDM has been somewhat frequently adopted by scholars in education to analyze educational data over the past few decades, especially after the 2000s; however, this practice remains relatively concentrated in the Western world (Baker & Yacef, 2009; Romero & Ventura, 2010). Some popular trends and topics for EDM researchers include clustering, relationship mining, prediction, a distillation of data for human judgment, and discovery with models (Baker & Yacef, 2009; Baker & Inventado, 2016; Romero & Ventura, 2010). Most cited papers using EDM investigate issues such as online courses (Zaïane, 2001), e-learning systems (Tang & McCalla, 2005; Zaïane, 2002), and student model development in their behavior, emotional, and engagement (Beck & Woolf, 2000; Baker & Yacef, 2009). Of the above topics, the discovery of models remains an emerging subarea of EDM (Baker & Yacef, 2009), especially when it is strategically combined with prediction. Admittedly, this method requires complex skills for writing algorithms relied upon by machine learning.

Thus far, only a limited number of research works employ machine learning as an application to further explore advanced models to predict student performance or knowledge using EDM results, partly due to the fact that large-scale, national representative educational datasets only started to appear more recently (Baker & Yacef, 2009), and the high threshold of algorithms being written for education major scholars (Romero & Ventura, 2010). In China, studies that utilize machine learning only started to emerge in the mid-1990s; even now, there is a dearth of educational research that adopt advanced methods such as machine learning (Yue & Xu, 2019). Currently, no existing study has employed both traditional statistical methods as well as machine learning to investigate students' second language performance. Moreover, no study has used machine learning as a supporting tool to evaluate the quality and effectiveness of the statistical models, which were built by sourcing from a combination of literature in English education, SLA, parental involvement, cognitive aptitude, and teacher attributes, among other research subjects.

In summary, existing literature on the related fields of student performance in China requires further research to identify effective predictors by using large educational datasets with advanced quantitative approaches. This study closely examined the CEPS dataset as well as the relationship between student English performance and its predictors using traditional statistical modeling and machine learning, with the intent to contribute to authentic educational theories that are specific to the context of China.

## Research Questions

The current study investigated students' English performance and its predictors at the individual level (student and parents), class level, and school level. Specifically, this study aims to answer the following research questions:

1. How are student variables at the individual level in the CEPS dataset associated with students' English performance?
2. How does teacher experience, along with other teacher attributes, contribute to student English performance?
3. How is parent involvement associated with student English performance?
4. How do school-level attributes (such as school SES) contribute to student English performance?
5. What portion of the variance in English performance can be explained by general cognitive aptitude?
6. Do machine learning processes generate models that better predict student English performance in the CEPS dataset? If so, how can these methods be used to improve educational research in China and other countries?

## Methods

Data

The China Education Panel Survey (CEPS) dataset is a large-scale, nationally representative, longitudinal survey that starts with two cohorts – the 7th and 9th graders in the 2013-2014 academic year. The collection of the CEPS data was funded by the following sources: 1) Renmin University Scientific Research Foundation; 2) Social Survey Foundation of National Survey Research Center in China; and 3) National Science Foundation (NSF) in the United States. The principal investigator is a professor at Remin University in China; the two co-PIs are professors at Johns Hopkins University and the University of Pennsylvania.

This panel survey consists of questionnaires from students, homeroom teachers, subject teachers (English, Chinese, and Mathematics), parents, and principals, which provides the platform and data for researchers worldwide to explore familial, social, cultural, and educational aspects of these populations. This dataset utilizes a stratified, multistage sampling design with probability proportional to size (PPS). A school-based sample of approximately 20,000 students was randomly selected from across 438 classrooms of 112 schools in 28 county-level units in Mainland China. The dataset started with the two cohorts, the 7th and 9th graders in the 2013-2014 academic year. The plan is to follow up with annual surveys in the 1st, 3rd, 7th, 8th, 17th, and 27th year after the students have graduated from junior high school. The CEPS website also published the 2014-2015 follow-up data from students who were 7th graders in the 2013-2014 baseline survey.

All questionnaires and data sourced from Chinese and English mediums were downloaded from the CEPS website (https://ceps.ruc.edu.cn/index.php?r=index/index&hl=en). The questionnaires (English and Chinese) were in PDF format, while all data were in the Stata format. In the 2013-2014 baseline survey, the student questionnaire contains 300 variables and 19,487 entries; the parent questionnaire contains 237 variables and 19,487 entries; the teacher questionnaire contains 853 variables and 438 entries; and the principal questionnaire contains 363 variables and 112 entries. The response rate for 2013-2014 was 98.7%. In the 2014-2015 follow-up survey for the 7th graders in 2013-2014, the student questionnaire contains 311 variables and 10,750 entries; the parent questionnaire contains 262 variables and 10,750 entries; the teacher questionnaire contains 179 variables and 791 entries; the principal questionnaire contains 304 variables and 112 entries. With a slight reduction from the 2013-2014 year, the response rate for the 2014-2015 wave was 91.9%. In this study, the following three data files were used to examine the students' English performance: 7th grade students in the first wave, 9th grade students in the first wave, and the 8th grade students in the second wave (who were the 7th graders from the previous year). These files comprehensively linked the student and the parent data (level 1, individual level), the English teacher data (level 2, class level), and the principal data (level 3, school level).

All data was publicly available on the CEPS website, and the multivariate nature of the data analysis in this paper helps to ensure the authenticity of the results presented below.

Variables

*Raw variables.* Tables 1 to 4 show the selected variables included from the original student questionnaire, the parent questionnaire, the English teacher questionnaire, and the principal questionnaire in both academic years. Other variables included are: school IDs, city IDs, and weights, for each wave.

**Table 1.**

*Student variables*

| **Questionnaire** | **Category** | **Question** | **Variable Name** | **Description** |
|---|---|---|---|---|
| Student | Personal background | / | sweight | Student individual weight |
| | | A1 | a01 | Sex |
| | | / | cog3pl | Student cognitive overall scores (with 3-IRT model) |
| | | / | stdchn | Midterm standardized Chinese score (fall semester) |
| | | / | stdeng | Midterm standardized English score (fall semester) |

| | | | | |
|---|---|---|---|---|
| | After-school academic-related activities | B14 | b14a1<br>b14a2<br>b14b1<br>b14b2 | Time spent on homework from school |
| | | B15 | b15a1<br>b15a2 | Time spent on homework from parents/cram school during weekdays |
| | | B16 | b16a1<br>b16a2 | Time spent on homework with parents/from cram school on weekends |
| | Outside tutor | B19 | b1904 | If English is being studied outside of school |
| | Perceptions of parents' expectations/ attitudes | B23 | b2301 | If parents care about homework and exam |
| | | B30 | b30 | Parents' requirements regarding academic ranking |
| | | B31 | b31 | Expectations for the future level of education achieved |
| | | B35 | b35 | If confident about their future |
| | | B32 | b32 | Feelings about parents' expectations |
| | Academic background | C7 | c07<br>c071 | If English has been studied in elementary school |
| | | C8 | c08 | Academic rank in Grade 6 |
| | | C12 | c12 | Academic rank at present |
| | Perceptions on English | C11 | c1103 | If English is perceived to be difficult at present |
| | | C13 | c1303<br>c1306<br>c1309 | English helps a lot for my future |
| | | | | My English teacher always asks me questions in class |

| | | | | |
|---|---|---|---|---|
| | | | | My English teacher always praises me |
| | Personal requirement | C22 | c22 | Highest education degree students expect to achieve |

**Table 2.**

*English teacher variables*

| **Questionnaire** | **Category** | **Question** | **Variable Name** | **Description** |
|---|---|---|---|---|
| English teacher | Instruction (other) | A7 | enga07 | Instruction hours |
| | | A8 | enga08 | Preparation hours |
| | Instructional approaches | A13 | enga1301 enga1302 enga1303 | Lectures |
| | | | | Group discussion |
| | | | | Interaction between teacher and students |
| | Instructional media | A14 | enga1401 enga1402 enga1403 enga1404 | Multimedia |
| | | | | Internet |
| | | | | Wall maps and other material |
| | | | | Personal teaching website |
| | Teaching materials | A15 | enga1501 enga1502 enga1503 enga1504 | Domestic textbook |
| | | | | Foreign textbooks |
| | | | | Reference materials |
| | | | | Materials designed by teachers |

| | | | | |
|---|---|---|---|---|
| | Personal background | B4 | engb04 | Educational diploma |
| | | B5 | engb05 | If I graduated from pedagogical universities |
| | | B6 | engb06 | If certified |
| | | B7 | engb07 | Years of teaching experience |
| | | B11 | engb11 engb11a | If institutionally registered teacher was admitted by the government |
| | | B12 | engb12 | Has professional title in teaching |

**Table 3.**

*Parent variables*

| Questionnaire | Category | Question | Variable Name | Description |
|---|---|---|---|---|
| Parent | Parental academic involvement | A4 | ba04 | If and how frequently do parents help with homework |
| | | A13 | ba13 | Time spent on child |
| | Family background | E7 | be07 | Parents' education background |
| | | E19 | be19 | Families' financial background |
| | Parent expectations | A18 | ba18 | The highest education level parents expect their children to achieve |
| | | C11 | bc11 | Parents' requirements for their children's academic rank |

**Table 4.**

*Principal variables*

| Questionnaire | Category | Question | Variable Name | Description |
|---|---|---|---|---|
| Principal | School information | A1 | pla01 | Category of school (e.g., public, private, etc.) |
| | | A23 | pla23 | School location (e.g., rural/urban, etc.) |
| | School population background information | B8 | plb08 | General parent education background |
| | | B9 | plb09 | General parent SES of school |

As shown in tables 1 to 4, four variables in the student data were test scores that were sourced directly from the sampled schools and the cognitive test administered. The first two scores were sourced from a cognitive aptitude test administered for this project. The CEPS Baseline Cognitive Ability Test Psychometric Report notes that this particular test measures the basic logical thinking and problem-solving skills of the subjects. It was designed as per the basic structure of the TEPS dataset in Taiwan to test the following 11 concepts in three dimensions: (1) language: analogy and verbal reasoning; (2) graphic and spatial understanding: graphic regularity analysis, origami questions, and geometric applications; (3) computational algorithms and reasoning: mathematical applications, custom operation rules, number sequence applications, abstract law analysis, probability, and numerical reverse thinking.

20 questions were used to cover the three dimensions for 7th graders while 22 questions were used for 9th graders. Both tests were to be completed within 15 minutes in class. The development team conducted two rounds of pilot testing and three rounds of revising. The Cronbach's Alpha for the 7th grade and 9$^{th}$ grade tests was 0.6892 and 0.7215, respectively. The first test score, "stcog", represents the number of cognitive test questions answered correctly by students – which determined the raw score. For 7th graders, this data ranged from 0 to 20. For 9th graders, the data ranged from 0 to 22. The second test score, "cog3pl", is the standardized score of the cognitive test calculated using the Three Parameter Item Response Theory (IRT) model, which is a commonly used paradigm for the design, analysis, and scoring of aptitude tests, based on the comparison and relationship between students' individual performance on the aptitude test and the overall ability the test is designed to measure. In the CEPS dataset, the researchers adopted the Three Parameter IRT (3-IRT), which accounts for the difficulty index, discriminative power index, and the guessing index (Baker & Kim, 2004). The three parameter estimates demonstrate that the probability of answering questions correctly just by guessing was very small (See Appendix I). The Pearson correlation matrix for the estimated Theta and raw scores was very large, indicating that the 3-IRT model was reliable in terms of scoring and analyzing the cognitive test (See Appendix II). The variable, "cog3pl", was selected for data analysis as it considers the different indices mentioned above. By analyzing the different

items on which the students scored correctly, this variable may appear differently across students even though some may have scored the same number of questions correctly.

The third and fourth test scores were midterm exam scores that were directly obtained from the schools. “stdchn” is the standardized midterm Chinese score in the fall semester of 2013-2014 for students’ native language (L1), while “stdeng” is the standardized midterm English score in the fall semester of 2013-2014 for students’ second language (L2). The standardized scores were calculated based on the different schools and grades, with the mean at 70 and a standard deviation of 10.

*Final variables.* Following the cleaning of data, a final list of variables is listed in Table 5.

**Table 5.**

*Final variables*

| Level of Data | Construct | Variable | Question |
|---|---|---|---|
| Level 1 - Individual level (Student) | Cognitive aptitude | cog3pl | Student cognitive overall scores (with 3-IRT model) |
| | Student performance | stdchn | Students’ Chinese score |
| | | stdeng | Students’ English score (outcome variable) |
| | Student personal background | sex | Gender |
| | Academic background | engoutside | If student has studied English in elementary school |

| | | | |
|---|---|---|---|
| | | rank_present | Academic rank at present |
| | Perceptions towards the English subject and English teacher | diff_eng_present | If they feel English is difficult at present |
| | | atti_eng_teacher | Students' attitudes towards their English teacher (composite score) |
| | Educational expectations | edulevel_exp | Highest education degree students expect to achieve |
| Level 1 - Individual level (Parent) | Parent involvement | help_with_hw | If and how frequently parents help with homework |
| | | time_on_child | Time spent on child |
| | Parent background | edu_background | Parents' education background |
| | | financial_background | Families' financial background |
| | Parent expectations for children's academic performance | p_edulevel_exp | Highest education level parents expect their children to achieve |
| | | p_req_rank | Parents' requirement for their children's academic rank |
| Level 2 - Class level (English teacher) | Instructional methods | instr_methods_lecture | Lectures |
| | | instr_methods_group | Group discussion |
| | | instr_methods_interact | Interaction between teacher and students |
| | Teaching facilities | facilities_multimedia | Multimedia |
| | | facilities_internet | Internet |
| | | facilities_posters | Wall maps and other materials |
| | | facilities_web | Personal teaching website |
| | Teaching | teaching_material | Teaching material English |

| | | | |
|---|---|---|---|
| | materials | | teacher is using (composite score) |
| | | teaching_material_foreign | If there's a foreign component in the teaching material |
| | Educational background | engt_edu_diploma | English teacher's educational diploma |
| | | engt_pedagogical_background | If graduated from pedagogical universities |
| | | engt_years_experience | Years of teaching experience |
| | | engt_government_registered | If institutionally registered teacher has been admitted by the government |
| | | engt_government_registered_year | Year when the English teacher registered with the government |
| Level 3 - School level (Principal) | School category | sch_category_public | If the school is public |
| | | sch_category_private | If the school is private |
| | | sch_category_pm | If the school is for migrant workers |
| | School location | sch_location_centercity | If the school is located in the center of the city |
| | | sch_location_outskirts | If the school is in the outskirts of the city |
| | | sch_location_ruralurban | If the school is in a rural-urban area |
| | | sch_location_rural | If the school is located in rural areas |
| | | sch_location_town | If the school is located in towns |
| | School-level parent education background | sch_parented | General parent education background |
| | School-level parent SES | sch_parentSES | General parent SES of school |

| | | | |
|---|---|---|---|
| Interaction effects | Cognitive aptitude score & current academic ranking | cog3pl:rank_present | Interaction between students' cognitive aptitude score and their academic rank |
| | Current academic ranking & educational expectations | rank_present:edulevel_exp | Interaction between students' current ranking in class and the highest education degree they expect to achieve |
| | Cognitive aptitude score & parents' requirement on ranking | cog3pl:p_req_rank | Interaction between students' cognitive aptitude scores and their parents' requirements on their academic ranking in class |
| | Educational expectations & parents' requirement on ranking | edulevel_exp:p_req_rank | Interaction between students' own educational level expectations and their parents' requirements for their academic ranking in class |

Data Cleaning

*Correlations.* Correlations among variables were measured to examine multicollinearity. For instance, consider a question in the parent survey that asks about the financial background of the family. There is another question in the principal survey that enquires about the general parent socioeconomic status of the school. Correlations between the two variables elicited from these questions are examined before maintaining both of them (correlations range from -0.01 to 0.21, indicating that these are not significant enough correlations and may be kept in the model).

*Coding.* The process of coding categorical variables from and to numeric variables was carried out using the *vlookup* function in Excel.

*Multiple Imputation.* Multiple imputation was used to manage the missing data for the variables in both waves (Young, Weckman, & Holland, 2011). Data cleaning for the three data files followed the same procedures to ensure consistency throughout the project. The data from the CEPS website were all taken in their numerically coded forms. In order to carry out multiple imputation with the MICE package in R, the numeric values were first coded back into their original categorical format. After this, the data was uploaded to R Studio Cloud, and the imputation method was specified for different types of data as follows: *Logreg* was used for binaries, *pmm* was used for continuous variables, *polyreg* was used for un-ordered multiple categories (more than 2), and *polr* was used for ordered variables with more than 2 categories. Five imputations (default) and 20 iterations were computed for each dataset. The iterations that were most similar (in means for continuous variables and frequencies for categorical variables) to the original dataset were selected for the final datasets. The datasets were then coded back into their numeric forms for data cleaning and subsequent analysis.

*Composite scores.* For continuous variables that enquired the same questions, for instance, those that asked for the number of hours and minutes a student spends on their homework; the response is converted into a single variable using the hour/minute ratio. Another example of the same would be the percentages of teaching material used by the English teacher. Composite scores were created using the percentage to decimal conversion, and a second variable identifying whether the teaching material contained any foreign component was also created.

*Multiple Component Analysis.* For categorical variables that measure the same latent construct, Multiple Component Analysis (MCA) was used to examine the contribution of each dimension and reduce the number of variables for each latent construct. To this end, *FactoMineR* and *factoextra* packages were used. The instructional methods and facilities used by the English teachers were expected to show a high correlation, and if confirmed, MCA would be used to reduce the number of variables as well as the dimensions. However, the three items for instructional methods and the four items for teaching facilities used were weakly correlated. The MCA results revealed multiple dimensions that required to be kept. Hence, the original items were kept in the final dataset.

*Using a single item as a predictor for some variables*. At present, there remain several debates regarding the use of single items as predictors in models across different disciplines (Adams et al., 1997; Fuchs & Diamantopoulos, 2009; Gardner et al., 1998). Multiple items measuring the same latent construct may capture more information to represent the construct well, while single items may be more practical. When grounded in theory, the single items may also capture the construct well. Some variables in this study are single-item questions. For instance, the teacher attribute items such as teachers' pedagogical background and their years of experience are relevant responses to single-item questions. Correlation matrices were computed and multiple component analysis was performed to decide if the various variables were strongly correlated or not, in order to determine whether the variables measured the same construct. Some single items were maintained as separate items as they did not measure the same latent construct as other questions but were of relevance to the purpose of the study.

*Centering.* Grand mean centering was applied to all level two (English teacher data) and level three variables (school principal data) to improve intercept interpretation. Centering the level two and level three variables makes the intercept for the outcome variable more interpretable as we consider the predictor at its mean rather than 0, which would not make any coherent sense.

*Linking Data.* To compile the final dataset for each grade in both waves, data from English teachers and principals were linked to student and parent data by matching student IDs, class IDs, school IDs, city IDs, and frames.

## Data Analysis

*Multilevel Modeling*. Multilevel modeling was used for statistical analysis to examine student English performance. RStudio and RStudio Cloud were used as the main platforms. The *lmer* function in the *lme4* package was used for multilevel modeling. Fixed effects were specified by adding the variables in the three levels of data to the models. Random effects that allow random intercepts were specified by adding the following expression to the models:

(1|schoolids) and (1|ctyids). The two-way interaction effects were specified by adding the product of two fixed effect variables to the models.

The general multilevel regression model is defined as below:

$$Y_i = \beta + S_iB_1 + C_iB_2 + SC_iB_3 + \varepsilon_1 + \varepsilon_2 + \varepsilon_3$$

where $Y_i$ represents the students' English performance for a student $i$,; $\beta$ is the constant, $S$ is a row vector of student-level variables, including the student and parent predictors; $B_1$ is a column vector of coefficients for the student level variables; $C$ is a row vector of class-level variables, which are the English teacher variables; $B_2$ is a column vector of coefficients for the class level variables; $SC$ is a row vector of school-level attributes; $B_3$ is a column vector of coefficients for the school level variables; and $\varepsilon$ is the residual term: $\varepsilon_1$ is the random error for cities, $\varepsilon_2$ is the random error for schools within cities, and $\varepsilon_3$ is the random error for students within schools and cities.

For all grade levels, models of levels 1, 2, and 3 were specified via the inclusion of variables listed in Table 5.

*Final Variables.* By examining the same group of students who were 7$^{th}$ graders in wave one and 8$^{th}$ graders in wave two, a final model was specified with the addition of the variable year in the fixed effects ("year") (random effects "year|ID" was dropped due to the small variability of the effect).

*Proportion of Variance Explained.* The *r.squaredGLMM* function in the *MuMIn* package was used to compute the R-squared information for the fitted model. Furthermore, the *ANOVA* function in the basic R package was used to compute the sum of the squared value for each fixed variable in each dataset before the sum of the squared value for each fixed variable was divided by the total sum of the squared value to obtain the proportion of variance explained by each fixed variable controlling for all other fixed variables in the model.

*Support Vector Regression*. Support vector machines can be used to predict a dichotomous response variable or a quantitative response variable. In case they are used to predict the latter, they are called support vector regression. In this study, a support vector machines regression model was used to predict the students' English performance scores. The algorithms in this context do not specify the model such as in the case of multilevel modeling. Instead, support vector regression learns from the entire dataset to generate an optimal model. Hence, the support vector regression model was not built by specifying the fixed and random variables based on existing theories and research studies, and subsequently, the variance explained by each fixed variable and the estimated slope for each fixed variable was not obtained from the mode. The *svm* function in the *e1071* package was used to perform support vector regression. Predictions were generated with the default radial kernel, the linear kernel, and the polynomial kernel (degree = 3). The *caret* package was used for k-fold cross-validation. The *Metrics* package was used to compute Mean Absolute Error (MAE) for accuracy purposes.

## Results

*1. How are student variables at the individual level in the CEPS dataset associated with students' English performance?*

Among the selected student variables of interest, Table 6 highlights those that are statistically significant, their estimated slopes, the 95% confidence intervals for the estimated slopes, and the proportion of variance explained. Interestingly, students' cognitive aptitude scores failed to explain the largest variance within the English achievement scores; however, they showed very large estimated slopes, indicating that one unit of change in the standardized cognitive aptitude scores (average min = -2.5, average max = 2.4) is associated with an average of more than four points of increase in their English achievement scores on a 1-100 scale (average 95% CI: [3.49, 4.91]). Other large estimated slopes were obtained in the students' current ranking in class (average 95% CI: [2.16, 2.92]); and whether they felt that English was a difficult language for them (average 95% CI: [3.86, 4.30]). The results indicated that one category of increase in students' rank in class (near the bottom, below average, about average, above average, and being one of the top 5) was associated with an estimated 2-point increase in their English scores; and one category of increase in "if they felt English was difficult for them" (very difficult, a bit difficult, not very difficult, and not difficult at all) was associated with about a 4-point increase in their English scores.

**Table 6.**

*Statistically significant student variables*

| Student Variables | Statistically significant in which dataset | Estimated slope | 95% confidence interval | t value | Proportion of variance explained |
|---|---|---|---|---|---|
| Cognitive aptitude score | 7th graders in wave one; 9th graders in wave one; 8th graders in wave two; 7th/8th graders in two waves | 2.58; 0.80; 5.60; 7.86 | [1.95, 3.16]; [0.20, 1.35]; [4.64, 6.57]; [7.17, 8.54] | 8.434; 2.723; 11.330; 22.454 | 19.40%; 18.73%; 34.13%; 22.40% |
| Chinese score | 7th graders in wave one; 9th graders in wave one; 8th graders in wave two; 7th/8th graders in two waves | 0.42; 0.38; 0.72; 0.79 | [0.41, 0.44]; [0.37, 0.40]; [0.69, 0.74]; [0.77, 0.81] | 53.584; 46.002; 56.610; 99.496 | 59.05%; 57.62%; 43.82%; 61.40% |
| Sex | 9th graders in wave one | 1.98 | [1.70, 2.25] | 14.042 | 1.70% |
| If they took English classes outside of school | 7th graders in wave one; 8th graders in wave two; 7th/8th graders in two waves | 0.54; 1.29; 1.66 | [0.21, 0.85]; [0.49, 2.09]; [1.15, 2.17] | 3.286; 3.162; 6.340 | 0.42%; 0.25%; 0.31% |
| Current academic ranking in class | 7th graders in wave one; 9th graders in wave one; | 3.04; 1.78 | [2.54, 3.54]; [1.27, 2.30] | 11.832; 6.752 | 11.20%; 13.14% |
| If English was difficult for them | 7th graders in wave one; 9th graders in wave one; 8th graders in wave two; 7th/8th graders in two waves | 2.41; 2.37; 7.74; 3.80 | [2.25, 2.55]; [2.20, 2.51]; [7.38, 8.10]; [3.59, 4.02] | 32.081; 30.009; 41.949; 35.069 | 7.34%; 5.72%; 16.87%; 7.23% |

| | | | | | |
|---|---|---|---|---|---|
| Their attitudes towards their English teacher | 7th graders in wave one; 8th graders in wave two; 7th/8th graders in two waves | 0.37; 1.76 2.90 | [0.19, 0.54]; [1.33, 2.19]; [2.62, 3.18] | 4.090; 7.949; 20.522 | 0.15%; 0.54%; 2.19% |
| Educational expectations | 7th graders in wave one; 9th graders in wave one; 8th graders in wave two; 7th/8th graders in two waves | 0.26; 0.32; 0.83; 0.60 | [0.02, 0.51]; [0.08, 0.56]; [0.33, 1.33]; [0.25, 0.95] | 2.127; 2.624; 3.262; 3.385 | 0.17%; 0.26%; 0.68%; 0.92% |

Cognitive aptitude scores, students' Chinese scores, whether or not English was difficult for students, and the students' own expectations for their educational achievement are four student variables that were statistically significant for all four datasets with a large proportion of variance explained as compared with other variables. Students' gender was only statistically significant for 9th graders in wave one, who were in their last year of middle school.

*2. How does teacher experience, along with other teacher attributes, contribute to student English performance?*

Most teacher variables were not consistently statistically significant across all datasets. On average, the teacher variables explained very little of the variance within the students' English achievement scores. The amount of variance and the t values for the statistically significant predictors in the different datasets served as a reference point to conclude whether the variables can be deemed as good predictors of English achievement scores as the effect sizes were small and not meaningfully different from the non-statistically significant results despite the fact that some variables were statistically significant. Table 7 summarizes the teacher variables that were statistically significant in the various datasets with estimated slopes, 95% confidence intervals, and the proportion of variance explained for further meaningful interpretation.

**Table 7.**

*Statistically significant English teacher variables*

| Teacher Variables | Statistically significant in which dataset | Estimated slope | 95% confidence interval | t value | Proportion of variance explained |
|---|---|---|---|---|---|
| Instructional methods – lecture | 7th/8th graders in two waves | 0.50 | [0.10, 0.90] | 2.442 | 0.01% |
| Instructional methods – group discussion | 7th/8th graders in two waves | -1.28 | [-1.62, -0.93] | -7.237 | 0.08% |
| Instructional methods – teacher/student interaction | 7th/8th graders in two waves | 1.06 | [0.68, 1.43] | 5.537 | 0.18% |
| Teaching facilities – multimedia | 9th graders in wave one; 8th graders in wave two | 0.41; 1.25 | [0.20, 0.57]; [0.51, 1.96]; | 4.319; 3.378 | 0.09%; 0.10% |
| Teaching facilities – internet | 9th graders in wave one; 7th/8th graders in two waves | 0.19; 0.32 | [0.03, 0.36]; [0.07, 0.57] | 2.282; 2.551 | 0.02%; 0.01% |

| | | | | | |
|---|---|---|---|---|---|
| Teaching facilities – posters | 7th graders in wave one; 7th/8th graders in two waves | 0.16; 0.50 | [0.00, 0.30]; [0.25, 0.74] | 2.123; 3.951 | 0.08%; 0.01% |
| Teaching facilities – web | 9th graders in wave one; 7th/8th graders in two waves | -0.20; -0.80 | [-0.38, -0.02]; [-1.08, -0.52] | -2.160; -5.589 | 0.01%; 0.21% |
| Teaching material | 7th graders in wave one; 8th graders in wave two | -1.04; -3.43 | [-1.55, -0.48]; [-5.66, -1.19] | -3.743; -2.991 | 0.12%; 0.04% |
| If the teaching material had a foreign component | 9th graders in wave one; 8th graders in wave two | -0.65; -1.41 | [-1.09, -0.20]; [-2.62, -0.15] | -2.826; -2.235 | 0.03%; 0.04% |
| English teacher’s diploma | 7th/8th graders in two waves | 0.53 | [0.18, 0.87] | 3.011 | 0.00% |
| If the English teacher had a pedagogical background | 8th graders in wave two | -2.73 | [-4.37, -1.04] | -3.188 | 0.02% |
| Certification types | 7th/8th graders in two waves | 1.89 | [0.64, 3.13] | 2.986 | 0.01% |
| The English teacher’s years of experience | 8th graders in wave two | -0.12 | [-0.19, -0.03] | -2.776 | 0.01% |
| If the English teacher was registered by the government | 7th graders in wave one; 9th graders in wave one; | -0.77; -0.92 | [-1.26, -0.19]; [-1.54, -0.27] | -2.762; -2.760 | 0.05%; 0.04% |
| English teachers’ title type | 9th graders in wave one; 8th graders in wave two; 7th/8th graders in two waves | 0.27; 1.45; 1.05 | [0.07, 0.44]; [0.82, 2.03]; [0.78, 1.33] | 2.832; 4.675; 7.416 | 0.05%; 0.15%; 0.32% |

*3. How is parent involvement associated with students' English performance?*

In the context of relevant parent variables, Table 8 demonstrates those that were statistically significant with estimated slopes, 95% confidence intervals, and the proportion of variance explained. "If parents helped with their children's homework" and parents' expectations for their children's educational achievement in the future are the two parent variables that were statistically significant for all four datasets. Family's financial background was only statistically significant for the 9th graders in wave one, who were in their last year in middle school and preparing for the high school entrance exam.

**Table 8.**

*Statistically significant parent variables*

| Parent Variables | Statistically significant in which dataset | Estimated slope | 95% confidence interval | t value | Proportion of variance explained |
|---|---|---|---|---|---|
| If parents helped with homework | 7th graders in wave one; 9th graders in wave one; 8th graders in wave two | -0.16; -0.30; -0.91; | [-0.27, -0.05]; [-0.43, -0.17]; [-1.18, -0.63] | -2.841; -4.533; -6.414; -5.741 | 0.03%; 0.14%; 0.24%; 0.09% |
| Time spent on their children | 7th graders in wave one; 9th graders in wave one; | -0.05; -0.05 | [-0.08, -0.01]; [-0.09, -0.00] | -2.666; -2.169 | 0.05%; 0.04% |
| Family's financial background | 9th graders in wave one; | -0.34 | [-0.57, 0.12] | -2.982 | 0.07% |
| Parents' requirement on their children's | 9th graders in wave one; 8th graders in wave two; | 1.51; -1.57; -1.09 | [0.89, 2.15]; [-2.95, -0.19]; [-1.99, -0.19] | 4.677; -2.222; -2.381 | 0.47%; 1.01%; 0.66% |

| | | | | | |
|---|---|---|---|---|---|
| academic ranking in their class | 7th/8th graders in two waves | | | | |
| Parents' expectations for their children's educational achievement in the future | 7th graders in wave one; 9th graders in wave one; 8th graders in wave two; 7th/8th graders in two waves | 0.18; 0.15; 0.89; 1.00 | [0.08, 0.27]; [0.05, 0.24]; [0.65, 1.13]; [0.85, 1.15] | 3.718; 2.985; 7.304; 12.916 | 0.09%; 0.12$; 0.64%; 0.84% |

*4. How do school level attributes (such as school SES) contribute to student English performance?*

None of the school attributes were consistently statistically significant across all datasets. The amounts of variance and t values for the statistically significant predictors in the different datasets served as a reference point to determine whether the variables are good predictors of English achievement scores. Even if certain variables were found to be statistically significant, the effect sizes that correspond with them were small and not meaningfully different from the non-statistically significant results. Table 9 summarizes some statistically significant school characteristics across different student groups with their estimated slopes, 95% confidence intervals, and the proportion of variance explained for more meaningful interpretation. If the school was private and for migrant workers only, and if the school was located in the center city - were two school attributes that were statistically significant for both 7th graders and 9th graders in wave one. If the school was located in the center city was also statistically significant upon comparing the English performance longitudinally for the 7th graders in wave one, who were later 8th graders in wave two, controlling for all other variables in the model. However, upon considering the proportion of variance explained by these variables, no variables were found to explain more than 2% of the total variance, indicating that these statistically significant variables may not be meaningful enough for explaining the difference in student English achievement scores.

**Table 9.**

*Statistically significant school variables*

| School Variables | Statistically significant in which dataset | Estimated slope | 95% confidence interval | t value | Proportion of variance explained |
|---|---|---|---|---|---|
| If the school was private | 9th graders in wave one | 6.06 | [3.00, 8.92] | 3.716 | 0.04% |
| If the school was private and for migrant workers only | 7th graders in wave one; 9th graders in wave one | 4.04; 4.06 | [1.78, 6.31]; [1.69, 6.37] | 3.262; 3.151 | 0.05%; 0.06% |
| If the school was located in the center city | 7th graders in wave one; 9th graders in wave one; 7th/8th graders in two waves | -0.99; -1.03; 3.58 | [-1.83, -0.18]; [-1.85, -0.15]; [1.54, 5.51] | -2.187; -2.219 3.828 | 0.12%; 0.08%; 0.17% |
| If the school was located in a township location | 7th/8th graders in two waves | -3.57 | [-4.70, -2.39] | -6.084 | 0.07% |
| School's overall parent education | 7th graders in wave one | -0.71 | [-1.10, -0.30] | -3.177 | 0.06% |

| | | | | | |
|---|---|---|---|---|---|
| background | | | | | |
| School's overall family socioeconomic status | 7th/8th graders in two waves | 2.38 | [1.64, 3.06] | 6.694 | 0.20% |

*5 What portion of the variance in English performance can be explained by general cognitive aptitude?*

Students' general cognitive aptitude is one of the few most significant contributors to their English performance. On average, general cognitive aptitude contributes to about one-fourth of the total variance in English performance, controlling for all other variables in the model. For the 7th graders in wave one, cognitive aptitude explained 19.40% of the total variance; for the 9th graders in wave one, cognitive aptitude explained 18.73%, for the 8th graders in wave two, it explained 34.13% of the total variance of English performance, for the longitudinal data for the 7th graders in wave one, who later represented the 8th graders in wave two, cognitive aptitude explained 22.40% of the total variance in their English performance. The average estimated slope for cognitive aptitude scores was also somewhat significant (4.21) on a 1-100 scale, indicating that one point of increase in students' standardized cognitive aptitude scores (ranging from -2.5 to 2.4) was associated with over 4 points of increase in students' English achievement scores on a 1-100 scale.

**Table 10.**

*Proportion of variance explained by each fixed variable*

| Variable | Proportion of variance | | | | |
|---|---|---|---|---|---|
| | 7th graders in wave 1 | 9th graders in wave 1 | 8th graders in wave 2 | 7th/8th graders | Average |
| Year | / | / | / | 0.74% | 0.74% |
| Statistical weight | 0.00% | 1.05% | 0.37% | 0.19% | 0.40% |
| Cognitive aptitude score | 19.40% | 18.73% | 34.13% | 22.40% | **23.67%** |
| Chinese score | 59.05% | 57.62% | 43.82% | 61.40% | **55.47%** |
| Gender | 0.00% | 1.70% | / | 0.00% | 0.57% |
| If they undertook English classes outside of school | 0.42% | 0.19% | 0.25% | 0.31% | 0.29% |
| Current academic ranking in class | 11.20% | 13.14% | / | / | **12.17%** |
| If the student felt English was difficult | 7.34% | 5.72% | 16.87% | 7.23% | **9.29%** |
| Attitudes toward English teacher | 0.15% | 0.03% | 0.54% | 2.19% | 0.73% |
| Educational achievement expectation | 0.17% | 0.26% | 0.68% | 0.92% | 0.51% |
| If parents helped with homework | 0.03% | 0.14% | 0.24% | 0.09% | 0.13% |
| Time spent on their children | 0.05% | 0.04% | / | / | 0.04% |
| Parents' educational background | 0.00% | 0.00% | / | / | 0.00% |
| Family's financial background | 0.01% | 0.07% | 0.03% | 0.01% | 0.03% |
| Parents' requirement for their children's ranking in class | 0.79% | 0.47% | 1.01% | 0.66% | 0.73% |
| Parents' requirement for their children's educational | 0.09% | 0.12% | 0.64% | 0.84% | 0.42% |

| | | | | | |
|---|---|---|---|---|---|
| degree in the future | | | | | |
| Instructional methods – lecture | 0.16% | 0.00% | 0.04% | 0.01% | 0.05% |
| Instructional methods – group discussion | 0.00% | 0.03% | 0.03% | 0.08% | 0.03% |
| Instructional methods – teacher/student interaction | 0.00% | 0.04% | 0.00% | 0.18% | 0.06% |
| Teaching facilities – multimedia | 0.00% | 0.09% | 0.10% | 0.05% | 0.06% |
| Teaching facilities – internet | 0.02% | 0.02% | 0.00% | 0.01% | 0.01% |
| Teaching facilities – posters | 0.08% | 0.01% | 0.02% | 0.01% | 0.03% |
| Teaching facilities – web/blogs | 0.01% | 0.01% | 0.06% | 0.21% | 0.07% |
| Teaching material | 0.12% | 0.00% | 0.04% | 0.00% | 0.04% |
| If teaching material has a foreign component | 0.02% | 0.03% | 0.04% | 0.01% | 0.02% |
| English teachers' diploma | 0.00% | 0.01% | 0.00% | 0.00% | 0.00% |
| English teachers' pedagogical background | 0.04% | 0.00% | 0.02% | 0.02% | 0.02% |
| English teacher's certification types | 0.04% | 0.01% | 0.06% | 0.01% | 0.03% |
| English teachers' years of experience | 0.02% | 0.00% | 0.01% | 0.18% | 0.05% |
| If the English teacher was registered by the government | 0.05% | 0.04% | 0.10% | 0.03% | 0.05% |
| English teachers' title types | 0.03% | 0.05% | 0.15% | 0.32% | 0.14% |
| School category – public | 0.01% | 0.01% | 0.00% | 0.01% | 0.01% |
| School category – private | 0.01% | 0.04% | / | / | 0.03% |
| School category – private and for migrant workers | 0.05% | 0.06% | 0.01% | 0.01% | 0.03% |
| School location – center city | 0.12% | 0.08% | 0.01% | 0.17% | 0.09% |
| School location – outskirts of city | 0.01% | 0.00% | 0.01% | 0.00% | 0.01% |

| | | | | | |
|---|---|---|---|---|---|
| School location – rural/urban | 0.01% | 0.02% | 0.00% | 0.04% | 0.02% |
| School location – townships | 0.00% | 0.00% | 0.00% | 0.07% | 0.02% |
| School's overall parent education background | 0.06% | 0.02% | 0.00% | 0.04% | 0.03% |
| School's overall parent financial background | 0.00% | 0.00% | 0.00% | 0.20% | 0.05% |
| Interaction effects between cognitive aptitude score and students' current ranking in class | 0.23% | 0.02% | / | / | 0.13% |
| Interaction effects between students' current ranking in class and their expectations regarding educational degree | 0.10% | 0.01% | / | / | 0.05% |
| Interaction effects between cognitive aptitude score and parents' requirement for their children's academic ranking in class | 0.07% | 0.02% | 0.10% | 1.22% | 0.35% |
| Interaction effects between students' expectations for educational degree and their parents' requirement for their educational degree | 0.02% | 0.04% | 0.01% | 0.00% | 0.02% |
| Total | 100.00% | 100.00% | 100.00% | 100% | 100.00% |

Considering the fixed effects, the few notable variables that explained the majority of variance in English performance in decreasing order are: students' Chinese score (an average of 54.47% of the total variance); cognitive aptitude score (an average of 23.67% of the total variance); students' current academic ranking in their class (an average of 12.17% of the total variance); and if the students felt English was difficult for them at the time of the survey (an average of 9.29% of the total variance). Table 10 lists the proportion of variance in students' English performance explained by each fixed variable across all datasets, controlling for all other variables in the model.

*6. Do machine learning processes generate models that better predict student English performance in the CEPS dataset? If so, how can these methods be used to improve educational research in China and other countries?*

After exploring the various machine learning algorithms, support vector machine (SVM), specifically, support vector regression (SVR), was chosen in consideration of the nature of analysis in this project, i.e., since the outcome variables were predicted to be quantitative and continuous. SVR generates predictions regarding the outcome variable by optimizing the models with the training set (set as 70% of the entire dataset for each grade level), followed by testing against the test set (the remaining 30% of the dataset). Two separate datasets for each grade level were fed into the SVR models: one with only the variables used for the multilevel modeling, and another with other variables that were originally included in the datasets but not selected due to their lack of theoretical significance in the multilevel models.

The efficiency of the predictions may be tested by computing the squared correlation coefficient between the predicted and the observed scores. This may also be carried out by comparing the mean absolute error (MAE) for the multilevel model and the SVR model. The table below lists data on the R squared and MAE for all datasets obtained from the SVR models using the radial, linear, and polynomial kernels. The radial and polynomial kernels (3-

degree) allow for non-linear relationships (e.g., quadratic and cubic relationships, and two-way and/or three-way interactions) while the radial kernel may perform much better than the polynomial kernel in terms of addressing the issue of overfitting. For the SVR models that utilized all variables in the datasets, the R squared and MAE returned by the polynomial kernel showed the best results; for the SVR models that used the same set of variables as the multilevel models, R squared and MAE returned by the three kernels demonstrated different results.

Table 11 shows the R squared and MAE returned by the models with only the variables used for the multilevel models and all-inclusive datasets. Table 12 shows the R squared and MAE in both the multilevel models as well as the SVR models with the highest R squared and least MAE (with the polynomial kernel for 7th graders and 9th graders in wave one, and the radial kernel for the 8th graders in wave two) using the same set of variables for comparison. These two tables demonstrate that the multilevel models explained 62.68% to 76.19% of the total variance for students' English performance, accounting for both random and fixed effects in the models. SVR models explained roughly similar amounts of variance for students' English performance; although this was lesser than the multilevel models that accounted for both fixed and random effects for 7th graders and 9th graders in wave one, with a slightly higher $R^2$ observed for 8th graders in wave two. Overall, the multilevel models and SVR models elicited similar results for R squared and MAE values, which indicates that the multilevel models did not miss the important predictors or relationships therein, since SVR models could capture it; furthermore, SVR models did not return a significantly larger amount of R-squared information. MAE results revealed that SVR models yielded consistent slightly better predictions of the students' English achievement scores as compared to the multilevel models.

**Table 11.**

*R-squared and MAE values for SVR models with the radial, linear, and polynomial kernels in the dataset with the same set of variables, multilevel modeling, and all-inclusive variables.*

| Dataset | | $R^2$ | | | MAE | | |
|---|---|---|---|---|---|---|---|
| | | **Radial** | **Linear** | **Polynomial** | **Radial** | **Linear** | **Polynomial** |
| 7th graders in wave one | All-inclusive dataset | 0.604 | 0.592 | **0.607** | 4.71 | 4.79 | **4.59** |
| | Same variables | **0.611** | 0.590 | 0.607 | **4.55** | 4.79 | 4.57 |
| 9th graders in wave one | All-inclusive dataset | 0.631 | 0.624 | **0.646** | 4.69 | 4.69 | **4.56** |
| | Same variables | **0.641** | 0.627 | 0.640 | **4.56** | 4.69 | 4.58 |
| 8th graders in wave two | All-inclusive dataset | 0.700 | 0.693 | **0.764** | 12.86 | 12.75 | **11.02** |
| | Same variables | 0.707 | 0.694 | **0.765** | 12.62 | 12.76 | **10.82** |

**Table 12.**

*R-squared and MAE for multilevel models and support vector regression models*

| Dataset | $R^2$ | | MAE | |
|---|---|---|---|---|
| | Multilevel Models | SVR | Multilevel Models | SVR |
| 7th graders in wave one | 0.627 | 0.611 | 4.66 | 4.55 |
| 9th graders in wave one | 0.660 | 0.641 | 4.56 | 4.56 |
| 8th graders in wave two | 0.762 | 0.765 | 11.19 | 10.82 |

For the longitudinal dataset with the same group of students who were $7^{th}$ graders in wave one and $8^{th}$ graders in wave two, a paired-sample t-test was performed to detect a statistically significant increase in their English scores ($p < .001$, $t = -9.802$, 95% CI: [2.11, 3.17]). The predictors that contributed the most variance of English performance were consistent with other datasets, including Chinese score (61.4%) with an estimated slope at 0.79 (95% CI: [0.77, 0.81]); cognitive aptitude score (22.4%) with an estimated slope at 7.86 (95% CI: [7.17, 8.54]); if they felt English was difficult for them (7.23%) with an estimated slope at 3.80 (95% CI: [3.59, 4.02]); their attitudes towards their English teacher (2.19%) with an estimated slope at 2.90 (95% CI: [2.62, 3.18]); their own educational achievement expectation (0.92%) with an estimated slope at 0.60 (95% CI: [0.25, 0.95]); their parents' requirement for their educational achievement in the future (0.84%) with an estimated slope at 1.00 (95% CI: [0.85, 1.15]); and their parents' requirement for their ranking in class (0.66%) with an estimated slope at -1.09 (95% CI: [-1.99, -0.19]). Of the above predictors, cognitive aptitude scores, if English was difficult for the students, and their attitudes toward their English teacher – all demonstrated greater estimated slopes with somewhat narrow confidence intervals, indicating them as strong predictors of English achievement scores within the dataset. Specifically, based on the estimated slopes and 95% confidence intervals, one point of change in their cognitive aptitude scores (min = -3.1, max = 2.3) was associated with 7.86 points increase in students' English achievement scores (95% CI: [7.17, 8.54]); one category of increase in if they felt English was difficult for them (very difficult, a bit difficult, not very difficult, and not difficult at all) was associated with an increase of about 3.80 points in English achievement scores (95% CI: [3.59, 3.18]); one category of increase in their attitudes towards their English teacher (if they strongly disagree, somewhat disagree, somewhat agree, and strongly agree with the following three statements: 1) English helps a lot with my future development; 2) My English teacher always asks me to answer questions in class; 3) My English teacher always praises me) was associated with an increase of 2.90 points in English achievement scores (95% CI: [2.62, 3.18]).

## Discussion

### What explains most of the variance in students' English performance?

*Students' Chinese scores explained most of the variance in their English performance*

Among all the predictors in the statistical analysis, students' Chinese score was found to have the largest t-statistics, explaining the most significant variance in their English performance across all four datasets (7th graders and 9th graders in wave one, 8th graders in wave two, and the longitudinal data for the 7th graders who were 8th graders in wave one and wave two, respectively).

This corroborates the literature on the influence of L1 on L2 (Jarvis & Pavlenko, 2008). Extant literature has adequately explained regarding the impact of L1 on L2 acquisition (Harrison & Kroll, 2007; Proctor et al., 2006). Current literature reveals that the stronger the students' L1 skills, the better their L2 performance will be (Proctor et al., 2006). The multilevel modeling analysis in this study showed consistent results, indicating that higher Chinese scores were associated with better English scores. However, there are more effective predictors of foreign language performance aside from students' native language. Empirical studies have revealed several important predictors of foreign language achievement, including but not limited to: one's linguistic background (native language), such as being monolingual or bilingual/multilingual (Maluch, et al., 2014), language anxiety (Aida, 1994), and motivation/attitude variables (Dörnyei, 1990). A few variables included in the analysis may be categorized into the abovementioned groups. For instance, students' attitudes toward their English teacher may be an indicator grouped as an attitude/motivation variable; similarly, if English was difficult for them at the time of the survey could be an indicator of students' language anxiety. The variance explained by each predictor may undergo a slight shift if these areas are systematically measured in the survey. With the limited number of variables and concepts/dimensions measured in the areas that may influence students' English performance, the impact of their Chinese score may have been highlighted due to the limited availability of other predictors within the dataset.

*Cognitive aptitude score is a good predictor of English performance*

Cognitive performance was found to be statistically significant for both the 7th graders and 9th graders in wave one, the 8th graders in wave two (who were the 7th graders in wave one), as well as the longitudinal examination of the 7th graders in wave one and 8th graders in wave two together. Cognitive performance was also considered to be one of the few good predictors for the students' English performance explained in terms of proportion of variance and estimated slopes alongside 95% confidence intervals.

As demonstrated above, the four predictors that explained most of the variance in students' English performance included: their Chinese score, cognitive aptitude score, current academic ranking in class, and if they felt English was difficult for them at the time of the survey. On average, the cognitive aptitude score explained 23.67% of the total variance in their English performance, which can be considered rather substantial. This result is consistent with existing literature that illustrates the positive relationship between general cognitive ability and student outcomes, and, in this particular case, students' English performance. Measured as one of the three main dimensions in the cognitive aptitude test, language aptitude is a key component of the test used for the CEPS dataset. Hence, the cognitive aptitude scores could reflect the students' language aptitude to a significant extent. Language aptitude, according to Caroll (1958, 1964), is defined as the rate of acquisition in the context of foreign languages. In consideration of literature that suggests language aptitude to be general across various languages (Carroll, 1964), the results from the current study prove that the students' linguistic ability was somewhat consistent across different languages – particularly Chinese (L1) and English (L2).

However, when switching angles, cognitive aptitude failed to explain all aspects of this phenomenon. Note that students' L1 had a much greater influence on their L2 performance, relative to the impact of cognitive aptitude scores. This corroborates the sociocultural SLA literature, which emphasizes the social factors and situational factors in terms of their influence on learners' language acquisition process in addition to the cognitive perspective. However, if it was possible to interview some of the students, stronger claims regarding their perceptions and reasonings for their English performance could have been made; however, with the quantitative data reported in this study, one may still conclude the potential existence of other social and situational factors that contributed to their English performance, such as if they felt English was difficult for them, attitudes towards their English teacher, and if they felt their English teacher liked them, praised them often, or paid attention to them.

## Patterns of Note

*9th graders showed a pattern different from other grades*

Families' financial background and the gender of students were only statistically significant to the context of 9$^{th}$ graders in wave one. Existing literature has documented how the socioeconomic status of a family may be associated with children's academic achievement (Caldas & Bankston, 1997; White, 1982). Despite the controversy surrounding the level of correspondence between the financial background of a family and students' academic achievement (White, 1982), this study found financial status not to be statistically significant except in the case of 9th graders, who were in their last year of middle school in the CEPS dataset. In this specific context, the relationship between SES and English achievement scores was negative, indicating that lower family SES was associated with better English achievement scores.

The Chinese context differs greatly from the Western world, where most educational literature is concentrated. Firstly, the testing system and the functioning of the school entrance screening are very different. For example, in the United States, no school at any level solely considers students' test scores as the only standard in recruiting or evaluating students. The advantages include an emphasis on the whole-child concept and also highlight the importance of how to reason, be with others, and build a profile that encompasses arts, sports, and community service, rather than solely on test-taking skills. However, in China, most schools continue to use students' entrance scores as the sole standard for acceptance into programs. Moreover, most schools tend to teach all middle school content within two years, and use the last year to review and practice. While this concern has remained controversial for decades, such a system allows students from lower SES families to scale the social ladder as they may be enrolled in top universities by working hard and practicing more. In China, the better the high schools/universities, the cheaper the tuition fees. For students from the lower end of the socioeconomic status spectrum to afford college or university, better universities are considered an ideal choice due to their affordability. By solely considering entrance exam scores rather than other profile areas such as arts and sports, the playing ground for students from lower SES backgrounds, who simply may not have the capability to build such profiles, is leveled. The results corresponding with the 9th graders indicate that students from lower socioeconomic backgrounds had much higher scores, which is indicative of their efforts towards attaining a higher rank in terms of academic grades, thereby securing their positions in good high schools and changing their lives for the better.

However, SES was not consistently statistically significant across grade levels, as it only explained less than 1% of the total variance in English achievement scores, on average indicating that it may not be one of the great predictors for English achievement despite having demonstrated statistical significance for 9th graders. This corroborates findings from extant literature on student achievement and SES in China (Butler, 2014). However, it is worth noting that while SES by itself may not be significantly indicative of student achievement in China, related variables such as parent expectations and direct/indirect behaviors, in addition to SES, may be associated with student achievement.

Relatedly, students' gender was only statistically significant for the 9th graders. This, in part, is aligned with existing Western and Chinese literature that documents the achievement gaps between girls and boys in China (Cole, 1997; Lai, 2010; Xu & Li, 2018). The fact that gender was only statistically significant for the students who were in their last year of middle school may be associated with the structure of the institution and the arrangement of instruction for the three years in middle school. Due to the entrance exam system, schools typically organize their instruction so that students may have enough time before the entrance exam to go over all the content learned throughout their middle school. Simply put, many schools tend to teach the entire curriculum of three years within two years, leaving the final year for a comprehensive review and repetitive practice to enhance test-taking skills and verify knowledge levels. Moreover, the test-oriented education system requires strong self-discipline and instruction-following qualities, wherein girls generally outperform boys (Duckworth & Seligman, 2006). However, results obtained on gender yielded interesting interpretations. On one hand, gender only contributes to less than 2% of the total variance in English achievement scores; on the other hand, the estimated slope for 9th graders was 1.98 (95% CI: [1.70, 2.25]), which is not a vividly significant value. These findings call for further research in this area to provide more fine-tuned statements with relatively more robust evidence. Existing studies that examine the gender difference/gender gap in the CEPS dataset mostly focused on overall academic achievement that consisted of Chinese, math, and English scores; and math achievement scores in particular. Gender did not have a large effect size in this study potentially because only the English achievement score was the targeted outcome variable, and because the same model may return different results in case math achievement scores or the overall academic achievement was being examined.

*Teacher-level and school-level variables showed small effect sizes*

Overall, teacher variables were not consistently statistically significant across all datasets. Specifically, teachers' years of experience were not statistically significant except in the case of 8th graders in wave two. This result may be further weakened by the small effect size via the examination of the t value and the proportion of variance explained (0.05%). The small amount of variance explained by teacher experience, despite its statistical significance in only one dataset, indicates that it would not be a good predictor for the middle school students' English achievement scores in this dataset. Existing literature, which is most concentrated in the western world, suggests that teacher attributes can influence the effectiveness of instruction that subsequently influences student performance. For example, expert and novice teachers show different patterns of instructional practices and tend to respond to student feedback differently (Borko & Livingston, 1989). However, no teacher variables in the current study were statistically significant for all grade levels. Moreover, on average, the teacher variables explained no larger than 1% of the total variance of their English performance.

Similarly, school-level variables also only explained very small proportions of variance (less than 1% for all school-level variables), indicating that they are not good predictors for students' English achievement scores despite their statistical significance in certain datasets, which may be explained partly by the large sample size.

This finding corresponds with how the "shortcut" (Wen & Xie, 2017) may not be the most suitable theoretical base in the Chinese context. For example, the variable of teacher background may explain further variance in the students' language performance; one of the reasons for this may be the large variability in teacher attributes owing to the different contexts of the country. Since the curriculum and instruction are more centralized in China and teachers are required to achieve certain goals to be qualified for teaching in various settings

- it restricts the variability of potential teacher attributes (such as certification or title) and their instruction, such as exemplified by curriculum/material use and instructional pacing. As a result, teacher variables may become non-statistically significant due to the small variability of the teacher variables.

Furthermore, teacher variables were statistically significant in the longitudinal examination of 7th graders in wave one, who became 8th graders in wave two, as compared to other student groups examined in this study. Similarly, these teacher variables generally explained no larger than 1% of the total variance of English achievement, indicating that they were not meaningful in interpreting English achievement despite their statistical significance.

In short, China needs its own, authentic educational theory to better depict and explain its educational context and localized concerns. Literature on how teacher attributes can contribute to students' English performance in the Chinese context is required to better interpret study implications and make better-informed policy decisions.

*Parents' expectations for their children's academic achievement.*

Among the variables that are consistently statistically significant and demonstrated somewhat large effect sizes in predicting the students' English performance, parents' expectations were a notable one. Apart from the few student variables that explained the most variance in their English performance, parent variables, which included if they helped with homework, time spent on their children, parents' requirement for their children's academic ranking in class, and parents' expectations on their children's educational achievement in the future, were statistically significant for at least two student groups among the 7th, 8th, and 9th graders in the CEPS dataset. Specifically, if they helped with their children's homework and their requirement for children's future educational achievement were statistically significant for all student groups in the dataset.

Parent expectations, in particular, are defined by Christenson et al. (1992) as future aspirations or current expectations for children's academic performance. It has been documented as a key predictor of student outcomes, perhaps much more so than family SES (White, 1982). In the Chinese context, wherein parents pay for bills, tuition, living expenses, and all other expenses related to study or extracurriculars, it is quite expected that parents' expectations generally play a significant role in their children's academic and non-academic outcomes. In this study sample, parent expectations - which are represented by two questions about the expectations of parents about 1) their children's academic ranking in their class; and 2) the educational degree they wished for their children to achieve in the future - explained the most variance in addition to the student individual predictors. This has great educational implications, as it indicates that parent expectations may play an important role in students' English performance than the class level or even school level predictors, which include but are not limited to their English teacher attributes, instruction, and school-wide SES. Simply put, this makes the students' English outcome more open to change and control as the variables explaining the most variance primarily remain on the family level: students' attitudes, perceptions, hard work, and parents' requirements and expectations, in contrast with uncontrollable areas such as English teacher attributes or school SES.

Taking all of the above factors and contexts into account, the bigger picture depicted by the results of this study sample explains how the students' English performance could be largely influenced by student variables, including their Chinese score (L1), their cognitive aptitude score, their responses to whether or not English was difficult for them, their attitudes towards their English teacher, and their educational achievement expectations; which is compounded by parent variables, including variables such as if their parents helped with their homework, their parents' expectations for academic ranking in classes, and parents' expectations for future educational achievement. While there were some other statistically significant variables in the class-level and school-level data, they only explained a very small proportion of the total variance in the students' English performance. However, further research is required to determine factors that may affect cognitive aptitude scores that were not included in the CEPS dataset, such as nutritional and social-emotional effects.

Interestingly enough, the best predictors of students' English performance analyzed in this study were mostly individual-level predictors, and variables such as their Chinese score, if English was difficult for them, students' attitudes towards their English teacher, and their own and parents' requirements/expectations are factors that may be changed or worked upon. Unlike predictors such as the socioeconomic status of the family and school-level attributes, which may not be easily changed or are outside of the family's own control, the good predictors revealed by this study (with the exception of students' cognitive aptitude scores) were aspects that they could work on to improve. This result is promising as students' English performance can be improved by working upon the few good predictors, which provide important pedagogical and social implications for teachers and school administrators of these students.

**The Role of Statistical Modeling and Machine Learning in Regression Analysis on Education Data**

*Multilevel models and SVR models explained similar amounts of variance in students' English performance*

The multilevel models obtained in this study were capable of explaining 60% to 76% of the total variance in the English performance of students across varying grade levels. SVR models that use all variables in the datasets and multilevel models with the same set of variables achieved similar results, with a slightly lower $R^2$ for the wave one students, and slightly better results for the wave two students. Given that the SVR models allow for non-linear relationships and interaction effects, the similar results (R-squared and MAE) elicited from the two different approaches indicated both models as successful and demonstrated that one particular model was not significantly better than the other. Note that MAE results for the SVR models were slightly smaller than the multilevel models in this study, indicating that the SVR methods could generate predictions with slightly better accuracy, albeit with very similar $R^2$ results. The differences were not meaningfully large when interpreting the results on a 1-100 scale for the English scores, although they may be of significance in other studies.

*Recommended utilization of statistical modeling and machine learning in regression analysis of education data*

Machine learning, represented by SVR in this paper, can be used as a reference for multilevel modeling in regression analysis for datasets such as observed in the study. The advantages of multilevel modeling include but are not limited to: 1) identifying specific predictors based on existing literature; and 2) the outputs of multilevel modeling return detailed results on the efficiency with which each fixed and random variable and/or factor can predict the outcome variable; for example, multilevel modeling enables researchers to obtain the estimated slopes in addition to confidence intervals, t values, and the proportion of variance explained by each fixed variable. In comparison, SVR algorithms do not elicit such detailed information, which would limit studies that use only the SVR approach to analyze educational datasets. However, since the SVR models allow for non-linear relationships such as quadratic effects, cubic effects, and/or two-way or three-way interactions, they permit researchers to test results against the statistical model and decide their efficiency. If the R-squared values returned by the machine learning method tend to be much larger than that of the statistical model, then it would imply that we may have missed important predictors or relationships. If SVR is not used alongside multilevel models, researchers would be unable to conclude that the most important predictors have been included within the human-specified model, because such a task would be improbable without a point of reference. Using the R-squared obtained from the SVR models to compare the efficiency of models in their interpretation of predictors for the students' English performance demonstrated that the use of both approaches may be significantly meaningful in interpreting the analysis results.

The fact that the proportion of variance could be explained by the model with human-specified variables and that the model using machine learning achieved similar results in this study indicated that the human-specified models have successfully included most of the important predictors for English achievement scores. This is because the machine learning methods, which allow for non-linear relationships and interaction effects, failed to return significantly better R-squared or MAE. However, such results may vary based on the datasets employed by the researcher. If the R-squared of the SVR model was significantly larger than that of the statistical model, it is probable that the researcher may have missed certain important predictors or relationships among the variables. In short, researchers will be able to make better-informed and confident conclusions on the efficiency of their human-specified models and predictors of their targeted outcome variable when both statistical methods and machine learning methods are used.

## Conclusions

This study examined predictors for students' English performance at the student/parent-level, class-level, and the school-level using the CEPS dataset. Results from the Multilevel modeling revealed that predictors explaining the most variance in English performance were individual-level predictors, including the students' Chinese score (L1), cognitive aptitude score, their perceptions of the difficulty of the language at the time of the survey, attitudes towards their English teacher, and expectations about their future educational achievement. Other good predictors include parent variables, such as parents' requirements and expectations for their children's ranking in class and future educational achievement. In short, student and parent variables explained the most variance in students' English performance. Aside from students' cognitive aptitude score, all other variables represent aspects that may be improved upon – this reveals that most good predictors of the students' English performance are controllable areas that families may work together to change, unlike in the case of financial capabilities, teacher quality, and school attributes.

Support vector regression with the radial kernel, the linear kernel, and the polynomial kernel with 10-fold cross-validation was used as a reference to predict the students' English performance. Similar R-squared values were obtained for the multilevel models and SVR models (both around 60-70%), indicating that by allowing for non-linear relationships and two-way and/or three-way interactions, the SVR models did not perform notably better than the multilevel models. In other words, the multilevel models successfully included most good predictors of English achievement scores. MAE results showed that SVR models performed consistently better in predicting the students' English achievement scores with smaller errors, although the difference was small on the 1-100 scale. Combining these two tangents, it can be concluded that SVR modeling may be used as a sufficient reference for statistical modeling to obtain a better understanding of the optimal model in the context of this dataset. This information is rather beneficial, as it indicates how the use of only one approach (either multilevel modeling or SVR) would result in losing out on data that would identify good predictors, estimated slopes, t statistics, confidence intervals, and proportions of variance explained by each fixed variable – it will also lead to losing the advantage of knowing what an optimal model could predict in terms of R-squared and MAE. Only by using both, with the machine learning approach as a reference, can researchers make more solid conclusions regarding the models to contribute further to the field.

## Limitations

One major limitation is that all relationships examined in this study are correlational, rather than causal. Hence, causal inferences may not be made while interpreting the association between the predictors and English achievement scores. Another limitation is that using the machine learning approaches would require a sufficiently large sample size, which would require more than a couple of hundred participants, which is a challenging task, particularly for educational researchers. Machine learning methods will not return good results with smaller sample sizes, which limits their use across various educational research projects, in addition to the relatively high threshold of computational skills required for algorithm writing.

Other limitations include the exclusion of variables such as the students' Hukou information. The original CEPS data included various aspects of the student such as academic-related questions, household-related questions, and social-emotional-related questions, among others. Another example of excluded variables is the potential factors that may affect a student's cognitive aptitude score (e.g., nutrition). Due to the limited scope of this dissertation, only education-related questions were selected to address the research questions. Future studies that examine the remaining variables may elicit valuable results.

Another limitation would be the limited data types included in the analysis. Though this study aims to examine the students' English achievement scores in the CEPS dataset using two different modeling methods, the inclusion of qualitative data such as focus groups, interviews, or case studies may have been more supportive in providing depth and meaning to the quantitative data obtained from the modeling analyses. For example, contacting some of the English teachers included in this data and obtaining their perspectives regarding their role and the impacts they believe they may have on their students would highlight further information and potential research questions to be addressed in this field. Moreover, only two waves of data were included in the analysis and discussion, while three out of the four datasets only examined the relationships between the variables for one year. More waves of data in the analysis would contribute to a more robust argument and would allow researchers to better understand the trends for one and multiple years for the same group of students. Future research should address this issue when more data becomes available on the CEPS official website.

## Implications

The results of this study revealed that students, parents, teachers, and school administrators may collaborate in an effort toward improving students' English performance. First, it is imperative that parents appropriately express their academic expectations to their children. Existing literature proves that parent involvement is positively associated with student achievement improvement. Moreover, parent expectations, defined by Christenson et al. (1992) as future aspirations or current expectations for children's academic performance, can greatly contribute to the total variance of student achievement (Christenson et al., 1992; White, 1982). Results of this study showed that parents' expectations for their children’s ranking in class as well as their future educational degree were found to be positively associated with their children’s English achievement scores. Additionally, parent expectations are one of the major factors that explain the total variance in English language achievement. If managed appropriately, a somewhat high expectation expressed from parents may facilitate improvements in English language achievement, controlling for other factors. However, note that there are several moving parts encompassed by these expectations; high expectations, if not managed appropriately, may result in further pressure and anxiety.

Secondly, parents may work together with their children to discuss children’s perceptions and attitudes towards the English subject and their English teacher. Social-emotional factors such as motivation and perceptions have been associated with students’ language learning process and achievement in the literature (Halliday, 1975, 1978, 1993; Krashen, 1982). In alignment with extant literature, results from the present study also showed that the more difficult students found English, the lower their English performance tended to be; the more they felt that their English teacher paid attention to them and praised them, the better their English scores would be. This allows for parents and children to collaboratively establish a healthy relationship with the English subject and English teacher, which may help to shape or reshape students' attitudes towards and perceptions of the subject and the teachers, consequently influencing their English achievement scores.

From the educator's perspective, English teachers must be encouraged to connect further with their students and pay attention to their emotional and psychological well-being in relation to their teaching of the English subject. As previously discussed, students' attitudes towards and perceptions of the English subject and their English teacher can contribute to a large portion of the total variance in their English achievement scores (Halliday, 1975, 1978, 1993; Krashen, 1982; Ortega, 2014). Chinese education research has largely not documented students' emotional and psychological wellbeing, and many teachers tend to focus mainly on academic matters and student achievement, while less attention is paid to the feelings of their students. Thus, teachers and school leaders should allocate more time and resources to address this matter.

Additionally, Chinese and English teachers may work together to strategize means by which they may improve scores in their respective language subjects. Researchers have shown that L1 can have a significant influence on language learners' learning process as well as the acquisition of foreign languages (Ortega, 2014). Results from the present study contribute additional evidence to the amount of total variance in English achievement scores explained by the students' Chinese performance, which, in this case, was their L1. Although the brainstorming of strategies would require further research, it is certain that students' Chinese and English performance were positively associated with each other, and if the teachers could work collaboratively, there may be a notable improvement in both subjects.

Methodologically speaking, future quantitative analysis of students' English performance may consider the use of statistical modeling as well as machine learning to predict student outcomes, granted that their sample size would permit it. Using both the R-squared and MAE for both the multilevel models and the machine learning models, I was able to conclude more confidently that the multilevel models I specified captured important predictors, as the SVR models, which allowed for non-linear relationships and two-way and/or three-way interactions, returned similar R-squared and MAE. Existing educational research using quantitative approaches to study student achievement typically uses only one of the two methods - either statistical modeling or machine learning. Results from this study revealed that using machine learning methods as a reference for statistical modeling can help researchers form a more comprehensive picture of an optimal model and determine the amount of total variance that can be explained by the given dataset. With only either of the two, there shall remain a missing link between the theory-driven, widely accepted proven predictors and the optimal model that allows for linear and non-linear relationships as well as two-way and/or three-way interactions. Only by utilizing the two can researchers make better-informed conclusions on the efficiency of their models. Although the two approaches may operate differently and demonstrate varying results across studies, the machine learning approaches can function as a reference for statistical modeling in terms of what an optimal model should achieve when the sample size is large enough for the machine learning methods to work well.

## Appendix I

### 3PL Model Parameters

| Grade 7 | | | | Grade 9 | | | |
|---|---|---|---|---|---|---|---|
| **#** | **Difficulty Index** | **Discriminative Power Index** | **Guessing Index** | **#** | **Difficulty Index** | **Discriminative Power Index** | **Guessing Index** |
| **1** | -0.86603 | 0.390882 | 0.087494 | **1** | -1.43294 | 0.431686 | 0.03916 |
| **2** | -0.21782 | 0.495177 | | **2** | -1.845 | 0.425234 | |
| **5** | -0.31216 | 0.793699 | | **3** | -1.43551 | 0.664734 | |
| **7** | 2.524089 | 0.686253 | | **4** | 1.030498 | 0.638315 | |
| **8** | 0.325089 | 1.26812 | | **5** | -0.92733 | 1.052449 | |
| **9** | -0.05998 | 1.458692 | | **7** | 0.401503 | 0.555581 | |
| **10** | -0.85159 | 0.994851 | | **8** | 1.084032 | 0.76834 | |
| **11** | -1.1188 | 0.9029 | | **9** | 1.331433 | 0.807445 | |
| **12** | -1.59877 | 1.268491 | | **10** | 0.352951 | 0.508029 | |
| **13** | -1.07292 | 2.18453 | | **12** | 0.465497 | 0.906184 | |
| **14** | 0.214053 | 0.891226 | | **13** | 0.586952 | 1.143454 | |
| **15** | 0.23759 | 1.746712 | | **14** | -0.04743 | 2.029005 | |
| **16** | -0.11884 | 2.53162 | | **15** | 1.003295 | 0.891305 | |
| **17** | 0.629049 | 1.477846 | | **16** | 0.692651 | 1.423485 | |
| **18** | 2.663288 | 0.661669 | | **17** | 1.167014 | 1.128959 | |
| **19** | 1.745291 | 1.563785 | | **18** | 1.029577 | 1.753127 | |
| **20** | 2.514116 | 0.689873 | | **19** | 1.671244 | 1.1827 | |
| | | | | **20** | 1.541128 | 1.281932 | |
| | | | | **21** | 1.753072 | 1.062701 | |
| | | | | **22** | 2.00573 | 1.618309 | |

## Appendix II

### Correlation Matrix for Raw Scores, 1PL, 2PL, and 3PL Models

| | Grade 7 | | | Grade 9 | | |
|---|---|---|---|---|---|---|
| | Raw Score | 1PL | 2PL | Raw Score | 1PL | 2PL |
| 1PL | 0.9997 | | | 0.9991 | | |
| 2PL | 0.9574 | 0.9575 | | 0.9662 | 0.967 | |
| 3PL | 0.9459 | 0.9458 | 0.9966 | 0.9559 | 0.9561 | 0.9951 |

## Appendix III

### Fixed and Random Effect Data Tables

***7th graders in wave 1***

Random effects for 7th graders in wave 1:

| Groups | Name | Variance | Std. Dev. |
|---|---|---|---|
| School IDs | (Intercept) | 1.0388 | 1.0192 |
| City IDs | (Intercept) | 0.8423 | 0.9178 |
| Residual | | 38.5161 | 6.2061 |

Fixed effects for 7th graders in wave 1:

| | Estimate | Std. Error | t value | Variance Explained |
|---|---|---|---|---|
| (Intercept) | 21.9407469 | 1.3949527 | 15.729 | |
| sweight | 0.0006536 | 0.0001779 | 3.675 | 0.00% |
| cognitive score | 2.5810654 | 0.3060216 | 8.434 | 19.40% |
| Chinese score | 0.4214368 | 0.0078649 | 53.584 | 59.05% |
| sex | 0.1267712 | 0.1225506 | 1.034 | 0.00% |
| time spent on homework | -0.0006833 | 0.0007210 | -0.948 | 0.00% |
| outside English class | 0.5395636 | 0.1641832 | 3.286 | 0.42% |
| current ranking | 3.0387562 | 0.2568166 | 11.832 | 11.20% |
| difficulty of English | 2.4106618 | 0.0751433 | 32.081 | 7.34% |
| attitudes towards English teacher | 0.3714519 | 0.0908291 | 4.090 | 0.15% |
| education level expectation | 0.2649009 | 0.1245640 | 2.127 | 0.17% |

| | | | | |
|---|---|---|---|---|
| help with homework | -0.1574884 | 0.0554409 | -2.841 | 0.03% |
| time on child | -0.0479844 | 0.0180008 | -2.666 | 0.05% |
| educational background | 0.0746142 | 0.0381314 | 1.957 | 0.00% |
| financial background | -0.0722967 | 0.1110768 | -0.651 | 0.01% |
| required ranking on child | 0.2611965 | 0.3209943 | 0.814 | 0.79% |
| education expectation on child | 0.1756248 | 0.0472300 | 3.718 | 0.09% |
| instructional methods_lecturing | -0.1251538 | 0.1300933 | -0.962 | 0.16% |
| instructional methods_group discussion | -0.0084603 | 0.1110798 | -0.076 | 0.00% |
| instructional methods_interaction | -0.0920080 | 0.1365563 | -0.674 | 0.00% |
| teaching facilities_multimedia | 0.0127138 | 0.1030862 | 0.123 | 0.00% |
| teaching facilities_internet | 0.0471272 | 0.0819669 | 0.575 | 0.02% |
| teaching facilities_posters | 0.1633634 | 0.0769671 | 2.123 | 0.08% |
| teaching facilities_web | -0.1186504 | 0.0867828 | -1.367 | 0.01% |
| teaching materials | -1.0379124 | 0.2772866 | -3.743 | 0.12% |
| teaching materials foreign | -0.2710610 | 0.2103289 | -1.289 | 0.02% |

| | | | | |
|---|---|---|---|---|
| diploma | 0.1178285 | 0.1084346 | 1.087 | 0.00% |
| pedagogical background | -0.0887552 | 0.2540900 | -0.349 | 0.04% |
| certification | 0.0319554 | 0.3602142 | 0.089 | 0.04% |
| years experience | 0.0115807 | 0.0119280 | 0.971 | 0.02% |
| government registered | -0.7693063 | 0.2753793 | -2.794 | 0.05% |
| title | 0.1804084 | 0.0970112 | 1.860 | 0.03% |
| school category_public | -0.0035877 | 0.7059140 | -0.005 | 0.01% |
| school category_private | 0.2279322 | 1.6742460 | 0.136 | 0.01% |
| school category_migrant workers | 4.0364292 | 1.2373243 | 3.262 | 0.05% |
| school location_center city | -0.9934480 | 0.4543266 | -2.187 | 0.12% |
| school location_outskirts | -0.3542623 | 0.5267267 | -0.673 | 0.01% |
| school location_rural urban | -0.2959831 | 0.4968062 | -0.596 | 0.01% |
| school location_town | 0.2215799 | 0.4303020 | 0.515 | 0.00% |
| school level parent education background | -0.7053839 | 0.2220607 | -3.177 | 0.06% |
| school parent SES | 0.1979868 | 0.2298996 | 0.861 | 0.00% |
| cognitive score * current ranking | -0.1974507 | 0.0725390 | -2.722 | 0.24% |

| | | | | |
|---|---|---|---|---|
| current rank * education level expectation | -0.1533913 | 0.0343664 | -4.463 | 0.10% |
| cognitive score * parent requirement on ranking | -0.3661634 | 0.0972324 | -3.766 | 0.07% |
| education level expectation * education expectation on child | 0.0830164 | 0.0437934 | 1.896 | 0.02% |

***9th graders in wave 1***

Random effects for 9th graders in wave 1:

| Groups | Name | Variance | Std. Dev. |
|---|---|---|---|
| School IDs | (Intercept) | 0.9498 | 0.9746 |
| City IDs | (Intercept) | 1.3055 | 1.1426 |
| Residual | | 35.354 | 5.9459 |

Fixed effects for 9th graders in wave 1:

| | Estimate | Std. Error | t value | Variance Explained |
|---|---|---|---|---|
| (Intercept) | 23.5149593 | 1.3673637 | 17.197 | |
| sweight | 0.0003681 | 0.0001763 | 2.088 | 1.05% |
| cognitive score | 0.8009062 | 0.2940988 | 2.723 | 18.73% |
| Chinese score | 0.3812972 | 0.0082888 | 46.002 | 57.62% |
| sex | 1.9768436 | 0.1407840 | 14.042 | 1.70% |
| time spent on homework | 0.0004646 | 0.0007146 | 0.650 | 0.04% |
| English tutor | 0.3111564 | 0.1703276 | 1.827 | 0.19% |

| | | | | |
|---|---|---|---|---|
| current ranking | 1.7823729 | 0.2639937 | 6.752 | 13.14% |
| difficulty of English | 2.3663148 | 0.0788524 | 30.009 | 5.72% |
| attitudes towards English teacher | 0.1743055 | 0.0911857 | 1.912 | 0.03% |
| education level expectation | 0.3177649 | 0.1211110 | 2.624 | 0.26% |
| help with homework | -0.2996448 | 0.0661052 | -4.533 | 0.14% |
| time on child | -0.0452883 | 0.0208758 | -2.169 | 0.04% |
| educational background | -0.0004969 | 0.0394229 | -0.013 | 0.00% |
| financial background | -0.3420246 | 0.1147098 | -2.982 | 0.07% |
| education expectation on child | 0.1476284 | 0.0494561 | 2.985 | 0.12% |
| required ranking on child | 1.5079704 | 0.3224110 | 4.677 | 0.47% |
| instructional methods_lecturi ng | -0.0398186 | 0.1154378 | -0.345 | 0.00% |
| instructional methods_group discussion | 0.0707834 | 0.1256895 | 0.563 | 0.03% |
| instructional methods_intera ction | -0.1647893 | 0.1264518 | -1.303 | 0.04% |
| teaching facilities_multim edia | 0.4090595 | 0.0947145 | 4.319 | 0.09% |
| teaching facilities_interne t | 0.1940128 | 0.0850055 | 2.282 | 0.02% |
| teaching facilities_poster | 0.0915892 | 0.0770710 | 1.188 | 0.01% |

| | | | | |
|---|---|---|---|---|
| s | | | | |
| teaching facilities_web | -0.2006533 | 0.0928937 | -2.160 | 0.01% |
| teaching materials | -0.0654662 | 0.3037278 | -0.216 | 0.00% |
| teaching materials foreign | -0.6522794 | 0.2308135 | -2.826 | 0.03% |
| diploma | -0.1147340 | 0.1074077 | -1.068 | 0.01% |
| pedagogical background | 0.3545323 | 0.3337470 | 1.062 | 0.00% |
| certification | -0.2707402 | 0.4329468 | -0.625 | 0.01% |
| years experience | -0.0119552 | 0.0132354 | -0.093 | 0.00% |
| government registered | -0.9202428 | 0.3331512 | -2.762 | 0.04% |
| title | 0.2709785 | 0.0956899 | 2.832 | 0.05% |
| school category_public | 0.6752009 | 0.6998732 | 0.965 | 0.01% |
| school category_private | 6.0591695 | 1.6307362 | 3.716 | 0.04% |
| school category_migrant workers | 4.0622830 | 1.2890068 | 3.151 | 0.06% |
| school location_center city | -1.0317765 | 0.4649324 | -2.219 | 0.08% |
| school location_outskirts | 0.1065868 | 0.5198676 | 0.205 | 0.00% |
| school location_rural urban | -0.6846704 | 0.5011457 | -1.366 | 0.02% |
| school location_town | 0.1426993 | 0.4276905 | 0.344 | 0.00% |

| | | | | |
|---|---|---|---|---|
| school level parent education background | -0.3284266 | 0.2191331 | -1.499 | 0.02% |
| school parent SES | -0.0434620 | 0.2248008 | -0.193 | 0.00% |
| cognitive score * education level expectation | 0.1221208 | 0.0743049 | 1.644 | 0.02% |
| rank_present * edulevel_exp | 0.0829336 | 0.0362784 | 2.286 | 0.01% |
| cognitive score * parent requirement on ranking | -0.0927512 | 0.0932633 | -0.995 | 0.02% |
| education level expectation * education expectation on child | -0.1160001 | 0.0449213 | -2.582 | 0.04% |

***8th graders in wave 2:***

Random effects for 8th graders in wave 2:

| Groups | Name | Variance | Std. Dev. |
|---|---|---|---|
| School IDs | (Intercept) | 50.51 | 7.107 |
| City IDs | (Intercept) | 39.33 | 6.271 |
| Residual | | 217.73 | 14.756 |

Fixed effects for 8th graders in wave 2:

| | Estimate | Std. Error | t value | Variance Explained |
|---|---|---|---|---|
| (Intercept) | -18.25 | 5.433 | -3.359 | |

| | | | | |
|---|---|---|---|---|
| wave 2 sweight | 0.002085 | 0.0008171 | 2.552 | |
| wave 1 wave 2 sweight | -0.002693 | 0.0008222 | -3.275 | 0.37% |
| cognitive score | 5.604 | 0.4946 | 11.330 | 34.13% |
| Chinese score | 0.7168 | 0.01266 | 56.610 | 43.82% |
| time spent on homework | 0.01984 | 0.003912 | 5.071 | 0.60% |
| English tutor | 1.289 | 0.4075 | 3.162 | 0.25% |
| difficulty of English | 7.739 | 0.1845 | 41.949 | 16.87% |
| attitudes towards English teacher | 1.760 | 0.2214 | 7.949 | 0.54% |
| education level expectation | 0.8318 | 0.2550 | 3.262 | 0.68% |
| help with homework | -0.9065 | 0.1413 | -6.414 | 0.24% |
| financial background | -0.4732 | 0.2725 | -1.737 | 0.03% |
| education expectation on child | 0.8908 | 0.1220 | 7.304 | 0.64% |
| required ranking on child | -1.567 | 0.7054 | -2.222 | 1.01% |
| instructional methods_lecturing | 0.4886 | 0.4198 | 1.164 | 0.04% |
| instructional methods_group discussion | -0.03666 | 0.4534 | -0.081 | 0.03% |
| instructional methods_interaction | 0.3859 | 0.4057 | 0.951 | 0.00% |
| teaching facilities_multimedia | 1.253 | 0.3709 | 3.378 | 0.10% |

| | | | | |
|---|---|---|---|---|
| teaching facilities_internet | 0.2961 | 0.2634 | 1.124 | 0.00% |
| teaching facilities_posters | -0.1637 | 0.3403 | -0.481 | 0.02% |
| teaching facilities_web | -0.6438 | 0.3484 | -1.848 | 0.06% |
| teaching materials | -3.432 | 1.147 | -2.991 | 0.04% |
| teaching materials foreign | -1.411 | 0.6313 | -2.235 | 0.04% |
| diploma | 0.3986 | 0.3858 | 1.033 | 0.00% |
| pedagogical background | -2.726 | 0.8552 | -3.188 | 0.02% |
| certification | 4.273 | 2.373 | 1.801 | 0.06% |
| years experience | -0.1156 | 0.04163 | -2.776 | 0.01% |
| government registered | 1.422 | 1.488 | 0.956 | 0.10% |
| title | 1.447 | 0.3095 | 4.675 | 0.15% |
| school category_public | -0.2363 | 4.273 | -0.055 | 0.00% |
| school category_migrant workers | -5.646 | 7.227 | -0.781 | 0.01% |
| school location_center city | 1.869 | 2.711 | 0.689 | 0.01% |
| school location_outskirts | 4.456 | 4.066 | 1.096 | 0.01% |
| school location_rural urban | -0.4635 | 2.929 | -0.158 | 0.00% |

| | | | | |
|---|---|---|---|---|
| school location_town | 0.8048 | 2.719 | 0.296 | 0.00% |
| school level parent education background | 0.3819 | 1.499 | 0.255 | 0.00% |
| school parent SES | 0.8590 | 1.445 | 0.594 | 0.00% |
| cognitive score * parent requirement on ranking | -0.6727 | 0.2082 | -3.231 | 0.10% |
| education level expectation * education expectation on child | -0.1177 | 0.1001 | -1.176 | 0.01% |

***7th graders/8th graders longitudinal:***

Random effects:

| Groups | Name | Variance | Std. Dev. |
|---|---|---|---|
| School IDs | (Intercept) | 19.96 | 4.468 |
| City IDs | (Intercept) | 7.01 | 2.648 |
| Residual | | 179.11 | 13.383 |

Fixed effects:

| | Estimate | Std. Error | t value | Variance Explained |
|---|---|---|---|---|
| (Intercept) | -7.0912908 | 3.1132696 | -2.278 | |
| year | -7.8109822 | 0.3630836 | -21.513 | 0.74% |
| sweight | 0.0003646 | 0.0004802 | 0.759 | 0.19% |
| cognitive score | 7.8602188 | 0.3500578 | 22.454 | 22.40% |
| Chinese score | 0.7898330 | 0.0079383 | 99.496 | 61.40% |

| | | | | |
|---|---|---|---|---|
| sex | 0.0648771 | 0.2001383 | 0.324 | 0.00% |
| time spent on homework | 0.0030402 | 0.0015135 | 2.009 | 0.15% |
| English tutor | 1.6627079 | 0.2622770 | 6.340 | 0.31% |
| difficulty of English | 3.8024167 | 0.1084281 | 35.069 | 7.23% |
| attitudes towards English teacher | 2.9021091 | 0.1414161 | 20.522 | 2.19% |
| education level expectation | 0.6012049 | 0.1775992 | 3.385 | 0.92% |
| help with homework | -0.5137965 | 0.0894897 | -5.741 | 0.09% |
| financial background | 0.2251863 | 0.1808768 | 1.245 | 0.01% |
| education expectation on child | 0.9984578 | 0.0773018 | 12.916 | 0.84% |
| required ranking on child | -1.0917495 | 0.4585445 | -2.381 | 0.66% |
| instructional methods_lecturing | 0.4969764 | 0.2035183 | 2.442 | 0.01% |
| instructional methods_group discussion | -1.2810091 | 0.1770007 | -7.237 | 0.08% |
| instructional methods_interaction | 1.0591424 | 0.1912795 | 5.537 | 0.18% |
| teaching facilities_multimedia | 0.1584242 | 0.1606148 | 0.986 | 0.05% |
| teaching facilities_internet | 0.3208567 | 0.1257980 | 2.551 | 0.01% |
| teaching facilities_poster | 0.4973408 | 0.1258656 | 3.951 | 0.01% |

| | | | | |
|---|---|---|---|---|
| s | | | | |
| teaching facilities_web | -0.7972107 | 0.1426278 | -5.589 | 0.21% |
| teaching materials | -0.0216039 | 0.4808570 | -0.045 | 0.00% |
| teaching materials foreign | 0.0378108 | 0.2886140 | 0.131 | 0.01% |
| diploma | 0.5297376 | 0.1759242 | 3.011 | 0.00% |
| pedagogical background | 0.6211302 | 0.4048691 | 1.534 | 0.02% |
| certification | 1.8944816 | 0.6343896 | 2.986 | 0.01% |
| years experience | 0.0267014 | 0.0181917 | 1.468 | 0.18% |
| government registered | -0.6136146 | 0.5158120 | -1.190 | 0.03% |
| title | 1.0543707 | 0.1421801 | 7.416 | 0.32% |
| school category_public | 1.6416515 | 2.4032932 | 0.683 | 0.01% |
| school category_migra nt workers | -2.8643641 | 4.2824748 | -0.669 | 0.01% |
| school location_center city | 3.5753465 | 0.9341051 | 3.828 | 0.17% |
| school location_outskirt s | -1.0624572 | 0.9061576 | -1.172 | 0.00% |
| school location_rural urban | 1.3720421 | 0.9346974 | 1.468 | 0.04% |
| school location_town | -3.5690146 | 0.5866010 | -6.084 | 0.07% |
| school level parent education | -0.1257583 | 0.3140202 | -0.400 | 0.04% |

| | | | | |
|---|---|---|---|---|
| background | | | | |
| school parent SES | 2.3775225 | 0.3551878 | 6.694 | 0.20% |
| cognitive score * parent requirement on ranking | -1.9486833 | 0.1207192 | -16.142 | 1.22% |
| education level expectation * education expectation on child | -0.0100942 | 0.0631308 | -0.160 | 0.00% |

## Appendix IV

## Multilevel Modeling R Code

### 7th graders in wave one:

```
library(readxl)
library(lme4)
library(MuMIn)
w1G7_imp_coded <- read_excel("/cloud/project/w1G7_imp_coded_final.xlsx",
                             sheet = "Sheet1",
                             col_names = TRUE)
final_model_1 <- lmer(stdeng ~ 1 + (1|schids) + (1|ctyids) + sweight + cog3pl
+ stdchn + sex + Time_on_hw_min + engoutside + rank_present + diff_eng_present
+ Atti_eng + edulevel_exp +
                        help_with_hw  +  time_on_child  +  edu_background  +
fincial_background + p_req_rank + p_edulevel_exp +
                        cog3pl*rank_present  +  rank_present*edulevel_exp  +
cog3pl*p_req_rank + p_req_rank*edulevel_exp +
                        instr_methods_lecture   +   instr_methods_group   +
instr_methods_interact  +  facilities_multimedia  +  facilities_internet  +
facilities_posters    +    facilities_web    +    teaching_material    +
teaching_material_foreign + engt_edu_diploma + engt_pedagogical_background +
engt_certification  +  engt_years_experience  +  engt_government_registered  +
engt_title +
                        sch_category_public   +   sch_category_private   +
sch_category_pm   +  sch_location_centercity   +  sch_location_outskirts  +
sch_location_ruralurban + sch_location_town + sch_parented + sch_parentSES,
data = w1G7_imp_coded)
final_model_1
summary(final_model_1)
anova(final_model_1)
r_squared_w1G7 <- r.squaredGLMM(object = final_model_1)
r_squared_w1G7
```

### 9th graders in wave one:

```
w1G9_imp_coded <- read_excel("/cloud/project/w1G9_imp_coded.xlsx",
                             sheet = "Sheet1",
                             col_names = TRUE)
final_model_2 <- lmer(stdeng ~ 1 + (1|schids) + (1|ctyids) + sweight + cog3pl
+  stdchn  +  sex  +  Time_on_hw_composite  +  engoutside  +  rank_present  +
diff_eng_present + Atti_eng + edulevel_exp +
                        help_with_hw  +  time_on_child  +  edu_background  +
financial_background + p_edulevel_exp + p_req_rank +
                        cog3pl*rank_present  +  rank_present*edulevel_exp  +
cog3pl*p_req_rank + p_req_rank*edulevel_exp +
                        instr_methods_lecture   +   instr_methods_group   +
instr_methods_interact  +  facilities_multimedia  +  facilities_internet  +
facilities_posters    +    facilities_web    +    teaching_material    +
```

```
teaching_material_foreign + engt_edu_diploma + engt_pedagogical_background +
engt_certification + engt_years_experience + engt_government_registered +
engt_title +
                        sch_category_public + sch_category_private +
sch_category_pm + sch_location_centercity + sch_location_outskirts +
sch_location_ruralurban + sch_location_town + sch_parented + sch_parentSES,
data = w1G9_imp_coded)
final_model_2
summary(final_model_2)
r_squared_w1G9 <- r.squaredGLMM(object = final_model_2)
r_squared_w1G9
anova(final_model_2)
```

## 8th graders in wave two:

```
w2G8_imp_coded <- read_excel("/cloud/project/w2G8_imp_coded.xlsx",
                             sheet = "Sheet1",
                             col_names = TRUE)
final_model_3 <- lmer(w2eng ~ 1 + (1|schids) + (1|ctyids) + w2sweight +
w1w2sweight + w2cog3pl + w2chn + Time_on_hw_min + engoutside + diff_eng_present
+ Atti_eng + edulevel_exp +
                        help_with_hw + financial_background + p_edulevel_exp
+ p_req_rank +
                        w2cog3pl*p_req_rank + p_req_rank*edulevel_exp +
                        eng_instr_methods_lecture + eng_instr_methods_group +
eng_instr_methods_interact + eng_facilities_multimedia +
eng_facilities_internet + eng_facilities_posters + eng_facilities_web +
teaching_material + teaching_material_foreign + eng_edu_diploma +
eng_pedagogical_background + eng_certification + eng_years_experience +
eng_government_registered + eng_title +
                        sch_category_public + sch_category_pm +
sch_location_centercity + sch_location_outskirts + sch_location_ruralurban +
sch_location_town + sch_parent_ed + sch_parentSES, data = w2G8_imp_coded)
final_model_3
summary(final_model_3)
r_squared_w2G8 <- r.squaredGLMM(object = final_model_3)
r_squared_w2G8
anova(final_model_3)
```

## 7th graders/8th graders longitudinal:

```
w1G7_w2G8_G7 <- read_excel("/cloud/project/w1G7_w2G8.xlsx",
                           sheet = "w1G7",
                           col_names = TRUE)
w1G7_w2G8_G8 <- read_excel("/cloud/project/w1G7_w2G8.xlsx",
                           sheet = "w2G8",
                           col_names = TRUE)
w1G7_w2G8_eng <- t.test(w1G7_w2G8_G7$stdeng, w1G7_w2G8_G8$w2eng, paired = TRUE,
alternative = "two.sided")
w1G7_w2G8_eng
w1G7_w2G8 <- read_excel("/cloud/project/w1G7_w2G8.xlsx",
                        sheet = "w1G7_w2G8",
                        col_names = TRUE)
```

```
final_model_4 <- lmer(stdeng ~ 1 + (1|schids) + (1|ctyids) + year + sweight +
cog3pl + stdchn + sex + Time_on_hw_min + engoutside + diff_eng_present +
Atti_eng + edulevel_exp +
                        help_with_hw + fincial_background + p_edulevel_exp +
p_req_rank +
                        cog3pl*p_req_rank + p_req_rank*edulevel_exp +
                        instr_methods_lecture + instr_methods_group +
instr_methods_interact + facilities_multimedia + facilities_internet +
facilities_posters + facilities_web + teaching_material +
teaching_material_foreign + engt_edu_diploma + engt_pedagogical_background +
engt_certification + engt_years_experience + engt_government_registered +
engt_title +
                        sch_category_public + sch_category_pm +
sch_location_centercity + sch_location_outskirts + sch_location_ruralurban +
sch_location_town + sch_parented + sch_parentSES, data = w1G7_w2G8)
final_model_4
summary(final_model_4)
r_squared_w1G7_w2G8 <- r.squaredGLMM(object = final_model_4)
r_squared_w1G7_w2G8
anova(final_model_4)
```

## Appendix V

### Support Vector Machine R Code

### Datasets with all-inclusive variables:

### 7th graders in wave one:

Radial kernel:

```
w1G7_imp_coded <- read_excel("/cloud/project/w1G7_imp_coded_final.xlsx",
                             sheet = "Sheet1",
                             col_names = TRUE)
index_w1G7 <- createDataPartition(w1G7_imp_coded$stdeng, p = .7, list =
FALSE, times = 1)
train_w1G7 <- w1G7_imp_coded[index_w1G7,]
test_w1G7 <- w1G7_imp_coded[-index_w1G7,]
ctrlspecs_w1G7 <- trainControl(method = "cv", number = 10)
cross_val_model <- train(stdeng ~ .,
                         data = train_w1G7,
                         method = "svmRadial",
                         preProcess = c("center", "scale"),
                         trControl = ctrlspecs_w1G7)
cross_val_model
cross_val_model$bestTune
tune_w1G7 <- expand.grid(
  C = c(0.25, 0.5, 1),
  sigma = c(0.001, 0.01, 0.1, 1)
)
cross_val_model_w1G7 <- train(stdeng ~ .,
                              data = train_w1G7,
                              method = "svmRadial",
                              trControl = ctrlspecs_w1G7,
                              preProcess = c("center", "scale"),
                              tuneGrid = tune_w1G7)
cross_val_model_w1G7
```

Linear kernel:

```
w1G7_imp_coded <- read_excel("/cloud/project/w1G7_imp_coded_final.xlsx",
                             sheet = "Sheet1",
                             col_names = TRUE)
index_w1G7 <- createDataPartition(w1G7_imp_coded$stdeng, p = .7, list =
FALSE, times = 1)
train_w1G7 <- w1G7_imp_coded[index_w1G7,]
test_w1G7 <- w1G7_imp_coded[-index_w1G7,]
ctrlspecs_w1G7 <- trainControl(method = "cv", number = 10)
cross_val_model <- train(stdeng ~ .,
                         data = train_w1G7,
                         method = "svmLinear",
                         preProcess = c("center", "scale"),
                         trControl = ctrlspecs_w1G7)
cross_val_model
```

Polynomial kernel:

```
w1G7_imp_coded <- read_excel("/cloud/project/w1G7_imp_coded_final.xlsx",
                             sheet = "Sheet1",
                             col_names = TRUE)
index_w1G7 <- createDataPartition(w1G7_imp_coded$stdeng, p = .7, list =
FALSE, times = 1)
train_w1G7 <- w1G7_imp_coded[index_w1G7,]
test_w1G7 <- w1G7_imp_coded[-index_w1G7,]
ctrlspecs_w1G7 <- trainControl(method = "cv", number = 10)
cross_val_model <- train(stdeng ~ .,
                         data = train_w1G7,
                         method = "svmPoly",
                         preProcess = c("center", "scale"),
                         trControl = ctrlspecs_w1G7)
cross_val_model
cross_val_model$bestTune
tune_w1G7 <- expand.grid(
  degree = 3,
  C = 1,
  scale = 0.01
)
cross_val_model_w1G7 <- train(stdeng ~ .,
                              data = train_w1G7,
                              method = "svmPoly",
                              trControl = ctrlspecs_w1G7,
                              preProcess = c("center", "scale"),
                              tuneGrid = tune_w1G7)
cross_val_model_w1G7
```

## 9th graders in wave one:

Radial kernel:

```
w1G9_imp_coded <- read_excel("/cloud/project/w1G9_imp_coded.xlsx",
                             sheet = "Sheet1",
                             col_names = TRUE)
index_w1G9 <- createDataPartition(w1G9_imp_coded$stdeng, p = .7, list =
FALSE, times = 1)
train_w1G9 <- w1G9_imp_coded[index_w1G9,]
test_w1G9 <- w1G9_imp_coded[-index_w1G9,]
ctrlspecs_w1G9 <- trainControl(method = "cv", number = 10)
cross_val_model_w1G9_1 <- train(stdeng ~ .,
                                data = train_w1G9,
                                method = "svmRadial",
                                preProcess = c("center", "scale"),
                                trControl = ctrlspecs_w1G9)
cross_val_model_w1G9_1
cross_val_model_w1G9_1$bestTune
tune_w1G9 <- expand.grid(
  C = c(0.25, 0.5, 1),
  sigma = c(0.001, 0.01, 0.1, 1)
)
cross_val_model_w1G9 <- train(stdeng ~ .,
                              data = train_w1G9,
                              method = "svmRadial",
                              trControl = ctrlspecs_w1G9,
```

```
                              preProcess = c("center", "scale"),
                              tuneGrid = tune_w1G9)
cross_val_model_w1G9
```

Linear kernel:

```
w1G9_imp_coded <- read_excel("/cloud/project/w1G9_imp_coded.xlsx",
                             sheet = "Sheet1",
                             col_names = TRUE)
index_w1G9 <- createDataPartition(w1G9_imp_coded$stdeng, p = .7, list =
FALSE, times = 1)
train_w1G9 <- w1G9_imp_coded[index_w1G9,]
test_w1G9 <- w1G9_imp_coded[-index_w1G9,]
ctrlspecs_w1G9 <- trainControl(method = "cv", number = 10)
cross_val_model_w1G9_1 <- train(stdeng ~ .,
                                data = train_w1G9,
                                method = "svmLinear",
                                preProcess = c("center", "scale"),
                                trControl = ctrlspecs_w1G9)
cross_val_model_w1G9_1
```

Polynomial kernel:

```
w1G9_imp_coded <- read_excel("/cloud/project/w1G9_imp_coded.xlsx",
                             sheet = "Sheet1",
                             col_names = TRUE)
index_w1G9 <- createDataPartition(w1G9_imp_coded$stdeng, p = .7, list =
FALSE, times = 1)
train_w1G9 <- w1G9_imp_coded[index_w1G9,]
test_w1G9 <- w1G9_imp_coded[-index_w1G9,]
ctrlspecs_w1G9 <- trainControl(method = "cv", number = 10)
cross_val_model_w1G9_1 <- train(stdeng ~ .,
                                data = train_w1G9,
                                method = "svmPoly",
                                preProcess = c("center", "scale"),
                                trControl = ctrlspecs_w1G9)
cross_val_model_w1G9_1
cross_val_model_w1G9_1$bestTune
tune_w1G9 <- expand.grid(
  degree = 3,
  C = 1,
  scale = 0.01
)
cross_val_model_w1G9 <- train(stdeng ~ .,
                              data = train_w1G9,
                              method = "svmPoly",
                              trControl = ctrlspecs_w1G9,
                              preProcess = c("center", "scale"),
                              tuneGrid = tune_w1G9)
cross_val_model_w1G9
```

## 8th graders in wave two:

Radial kernel:

```
w2G8_imp_coded <- read_excel("/cloud/project/w2G8_imp_coded.xlsx",
```

```
                             sheet = "Sheet1",
                             col_names = TRUE)
index_w2G8 <- createDataPartition(w2G8_imp_coded$w2eng, p = .7, list =
FALSE, times = 1)
train_w2G8 <- w2G8_imp_coded[index_w2G8,]
test_w2G8 <- w2G8_imp_coded[-index_w2G8,]
ctrlspecs_w2G8 <- trainControl(method = "cv", number = 10)
cross_val_model_w2G8_1 <- train(w2eng ~ .,
                                data = train_w2G8,
                                method = "svmRadial",
                                preProcess = c("center", "scale"),
                                trControl = ctrlspecs_w2G8)
cross_val_model_w2G8_1
cross_val_model_w2G8_1$bestTune
tune_w2G8 <- expand.grid(
  C = c(0.25, 0.5, 1),
  sigma = c(0.001, 0.01, 0.1, 1)
)
cross_val_model_w2G8 <- train(w2eng ~ .,
                              data = train_w2G8,
                              method = "svmRadial",
                              trControl = ctrlspecs_w2G8,
                              preProcess = c("center", "scale"),
                              tuneGrid = tune_w2G8)
cross_val_model_w2G8
```

Linear kernel:

```
w2G8_imp_coded <- read_excel("/cloud/project/w2G8_imp_coded.xlsx",
                             sheet = "Sheet1",
                             col_names = TRUE)
index_w2G8 <- createDataPartition(w2G8_imp_coded$w2eng, p = .7, list =
FALSE, times = 1)
train_w2G8 <- w2G8_imp_coded[index_w2G8,]
test_w2G8 <- w2G8_imp_coded[-index_w2G8,]
ctrlspecs_w2G8 <- trainControl(method = "cv", number = 10)
cross_val_model_w2G8_1 <- train(w2eng ~ .,
                                data = train_w2G8,
                                method = "svmLinear",
                                preProcess = c("center", "scale"),
                                trControl = ctrlspecs_w2G8)
cross_val_model_w2G8_1
```

Polynomial kernel:

```
w2G8_imp_coded <- read_excel("/cloud/project/w2G8_imp_coded.xlsx",
                             sheet = "Sheet1",
                             col_names = TRUE)
index_w2G8 <- createDataPartition(w2G8_imp_coded$w2eng, p = .7, list =
FALSE, times = 1)
train_w2G8 <- w2G8_imp_coded[index_w2G8,]
test_w2G8 <- w2G8_imp_coded[-index_w2G8,]
ctrlspecs_w2G8 <- trainControl(method = "cv", number = 10)
cross_val_model_w2G8_1 <- train(w2eng ~ .,
                                data = train_w2G8,
                                method = "svmPoly",
                                preProcess = c("center", "scale"),
                                trControl = ctrlspecs_w2G8)
cross_val_model_w2G8_1
cross_val_model_w2G8_1$bestTune
tune_w2G8 <- expand.grid(
```

```
  degree = 2,
  C = 0.5,
  scale = 0.1
)
cross_val_model_w2G8 <- train(w2eng ~ .,
                              data = train_w2G8,
                              method = "svmPoly",
                              trControl = ctrlspecs_w2G8,
                              preProcess = c("center", "scale"),
                              tuneGrid = tune_w2G8)
cross_val_model_w2G8
```

### Datasets of the same set of variables with the multilevel models:

### 7th graders in wave one:

Radial kernel:

```
w1G7_imp_coded_v2 <- read_excel("w1G7_imp_coded_final.xlsx",
                                sheet = "For SVR",
                                col_names = TRUE)
index_w1G7_v2 <- createDataPartition(w1G7_imp_coded_v2$stdeng, p = .7, list
= FALSE, times = 1)
train_w1G7_v2 <- w1G7_imp_coded_v2[index_w1G7_v2,]
test_w1G7_v2 <- w1G7_imp_coded_v2[-index_w1G7_v2,]
ctrlspecs_w1G7_v2 <- trainControl(method = "cv", number = 10)
cross_val_model_v2 <- train(stdeng ~ .,
                            data = train_w1G7_v2,
                            method = "svmPoly",
                            preProcess = c("center", "scale"),
                            trControl = ctrlspecs_w1G7_v2)
cross_val_model_v2
cross_val_model_v2$bestTune
tune_w1G7_v2 <- expand.grid(
  degree = 3,
  C = 0.25,
  scale = 0.01
)
cross_val_model_w1G7_V2_2 <- train(stdeng ~ .,
                                   data = train_w1G7_v2,
                                   method = "svmPoly",
                                   trControl = ctrlspecs_w1G7_v2,
                                   preProcess = c("center", "scale"),
                                   tuneGrid = tune_w1G7_v2)
cross_val_model_w1G7_V2_2
```

Linear kernel:

```
w1G7_imp_coded_v2 <- read_excel("w1G7_imp_coded_final.xlsx",
                                sheet = "For SVR",
                                col_names = TRUE)
index_w1G7_v2 <- createDataPartition(w1G7_imp_coded_v2$stdeng, p = .7, list
= FALSE, times = 1)
train_w1G7_v2 <- w1G7_imp_coded_v2[index_w1G7_v2,]
test_w1G7_v2 <- w1G7_imp_coded_v2[-index_w1G7_v2,]
```

```
ctrlspecs_w1G7_v2 <- trainControl(method = "cv", number = 10)
cross_val_model_v2 <- train(stdeng ~ .,
                            data = train_w1G7_v2,
                            method = "svmLinear",
                            preProcess = c("center", "scale"),
                            trControl = ctrlspecs_w1G7_v2)
cross_val_model_v2
```

Polynomial kernel:

```
w1G7_imp_coded_v2 <- read_excel("w1G7_imp_coded_final.xlsx",
                                sheet = "For SVR",
                                col_names = TRUE)
index_w1G7_v2 <- createDataPartition(w1G7_imp_coded_v2$stdeng, p = .7, list
= FALSE, times = 1)
train_w1G7_v2 <- w1G7_imp_coded_v2[index_w1G7_v2,]
test_w1G7_v2 <- w1G7_imp_coded_v2[-index_w1G7_v2,]
ctrlspecs_w1G7_v2 <- trainControl(method = "cv", number = 10)
cross_val_model_v2 <- train(stdeng ~ .,
                            data = train_w1G7_v2,
                            method = "svmRadial",
                            preProcess = c("center", "scale"),
                            trControl = ctrlspecs_w1G7_v2)
cross_val_model_v2
cross_val_model_v2$bestTune
tune_w1G7_v2 <- expand.grid(
  C = c(0.25, 0.5, 1),
  sigma = c(0.001, 0.01, 0.1, 1)
)
cross_val_model_w1G7_V2_2 <- train(stdeng ~ .,
                                   data = train_w1G7_v2,
                                   method = "svmRadial",
                                   trControl = ctrlspecs_w1G7_v2,
                                   preProcess = c("center", "scale"),
                                   tuneGrid = tune_w1G7_v2)
cross_val_model_w1G7_V2_2
```

## 9th graders in wave one:

Radial kernel:

```
w1G9_imp_coded_v2 <- read_excel("w1G9_imp_coded.xlsx",
                                sheet = "For SVR",
                                col_names = TRUE)
index_w1G9_v2 <- createDataPartition(w1G9_imp_coded_v2$stdeng, p = .7, list
= FALSE, times = 1)
train_w1G9_v2 <- w1G9_imp_coded_v2[index_w1G9_v2,]
test_w1G9_v2 <- w1G9_imp_coded_v2[-index_w1G9_v2,]
ctrlspecs_w1G9_v2 <- trainControl(method = "cv", number = 10)
cross_val_model_w1G9_v2 <- train(stdeng ~ .,
                                 data = train_w1G9_v2,
                                 method = "svmRadial",
                                 preProcess = c("center", "scale"),
                                 trControl = ctrlspecs_w1G9_v2)
cross_val_model_w1G9_v2
cross_val_model_w1G9_v2$bestTune
tune_w1G9_v2 <- expand.grid(
```

```
  C = c(0.25, 0.5, 1),
  sigma = c(0.001, 0.01, 0.1, 1)
)
cross_val_model_w1G9_v2_2 <- train(stdeng ~ .,
                                   data = train_w1G9_v2,
                                   method = "svmRadial",
                                   trControl = ctrlspecs_w1G9_v2,
                                   preProcess = c("center", "scale"),
                                   tuneGrid = tune_w1G9_v2)
cross_val_model_w1G9_v2_2
```

Linear kernel:

```
w1G9_imp_coded <- read_excel("/cloud/project/w1G9_imp_coded.xlsx",
                             sheet = "Sheet1",
                             col_names = TRUE)
index_w1G9 <- createDataPartition(w1G9_imp_coded$stdeng, p = .7, list =
FALSE, times = 1)
train_w1G9 <- w1G9_imp_coded[index_w1G9,]
test_w1G9 <- w1G9_imp_coded[-index_w1G9,]
ctrlspecs_w1G9 <- trainControl(method = "cv", number = 10)
cross_val_model_w1G9_v2 <- train(stdeng ~ .,
                                 data = train_w1G9_v2,
                                 method = "svmLinear",
                                 preProcess = c("center", "scale"),
                                 trControl = ctrlspecs_w1G9_v2)
cross_val_model_w1G9_v2
```

Polynomial kernel:

```
w1G9_imp_coded_v2 <- read_excel("w1G9_imp_coded.xlsx",
                                sheet = "For SVR",
                                col_names = TRUE)
index_w1G9_v2 <- createDataPartition(w1G9_imp_coded_v2$stdeng, p = .7, list
= FALSE, times = 1)
train_w1G9_v2 <- w1G9_imp_coded_v2[index_w1G9_v2,]
test_w1G9_v2 <- w1G9_imp_coded_v2[-index_w1G9_v2,]
ctrlspecs_w1G9_v2 <- trainControl(method = "cv", number = 10)
cross_val_model_w1G9_v2 <- train(stdeng ~ .,
                                 data = train_w1G9_v2,
                                 method = "svmPoly",
                                 preProcess = c("center", "scale"),
                                 trControl = ctrlspecs_w1G9_v2)
cross_val_model_w1G9_v2
cross_val_model_w1G9_v2$bestTune
tune_w1G9_v2 <- expand.grid(
  degree = 3,
  C = 0.25,
  scale = 0.01
)
cross_val_model_w1G9_v2_2 <- train(stdeng ~ .,
                                   data = train_w1G9_v2,
                                   method = "svmPoly",
                                   trControl = ctrlspecs_w1G9_v2,
                                   preProcess = c("center", "scale"),
                                   tuneGrid = tune_w1G9_v2)
cross_val_model_w1G9_v2_2
```

## 8th graders in wave two:

### Radial kernel:

```
w2G8_imp_coded_v2 <- read_excel("w2G8_imp_coded.xlsx",
                                sheet = "For SVR",
                                col_names = TRUE)
index_w2G8_v2 <- createDataPartition(w2G8_imp_coded_v2$w2eng, p = .7, list =
FALSE, times = 1)
train_w2G8_v2 <- w2G8_imp_coded_v2[index_w2G8_v2,]
test_w2G8_v2 <- w2G8_imp_coded_v2[-index_w2G8_v2,]
ctrlspecs_w2G8_v2 <- trainControl(method = "cv", number = 10)
cross_val_w2G8_v2 <- train(w2eng ~ .,
                           data = train_w2G8_v2,
                           method = "svmRadial",
                           preProcess = c("center", "scale"),
                           trControl = ctrlspecs_w2G8_v2)
cross_val_w2G8_v2
cross_val_w2G8_v2$bestTune
tune_w2G8_v2 <- expand.grid(
  C = c(0.25, 0.5, 1),
  sigma = c(0.001, 0.01558857, 0.1, 1)
)
cross_val_model_w2G8_v2 <- train(w2eng ~ .,
                                 data = train_w2G8_v2,
                                 method = "svmRadial",
                                 preProcess = c("center", "scale"),
                                 trControl = ctrlspecs_w2G8_v2,
                                 tuneGrid = tune_w2G8_v2)
cross_val_model_w2G8_v2
```

### Linear kernel:

```
w2G8_imp_coded_v2 <- read_excel("w2G8_imp_coded.xlsx",
                                sheet = "For SVR",
                                col_names = TRUE)
index_w2G8_v2 <- createDataPartition(w2G8_imp_coded_v2$w2eng, p = .7, list =
FALSE, times = 1)
train_w2G8_v2 <- w2G8_imp_coded_v2[index_w2G8_v2,]
test_w2G8_v2 <- w2G8_imp_coded_v2[-index_w2G8_v2,]
ctrlspecs_w2G8_v2 <- trainControl(method = "cv", number = 10)
cross_val_w2G8_v2 <- train(w2eng ~ .,
                           data = train_w2G8_v2,
                           method = "svmLinear",
                           preProcess = c("center", "scale"),
                           trControl = ctrlspecs_w2G8_v2)
cross_val_w2G8_v2
```

### Polynomial kernel:

```
w2G8_imp_coded_v2 <- read_excel("w2G8_imp_coded.xlsx",
                                sheet = "For SVR",
                                col_names = TRUE)
index_w2G8_v2 <- createDataPartition(w2G8_imp_coded_v2$w2eng, p = .7, list =
FALSE, times = 1)
train_w2G8_v2 <- w2G8_imp_coded_v2[index_w2G8_v2,]
test_w2G8_v2 <- w2G8_imp_coded_v2[-index_w2G8_v2,]
ctrlspecs_w2G8_v2 <- trainControl(method = "cv", number = 10)
cross_val_w2G8_v2 <- train(w2eng ~ .,
                           data = train_w2G8_v2,
```

```
                           method = "svmPoly",
                           preProcess = c("center", "scale"),
                           trControl = ctrlspecs_w2G8_v2)
cross_val_w2G8_v2
cross_val_w2G8_v2$bestTune
tune_w2G8_v2 <- expand.grid(
  degree = 3,
  scale = 0.1,
  C = 0.25
)
cross_val_model_w2G8_v2 <- train(w2eng ~ .,
                                 data = train_w2G8_v2,
                                 method = "svmPoly",
                                 preProcess = c("center", "scale"),
                                 trControl = ctrlspecs_w2G8_v2,
                                 tuneGrid = tune_w2G8_v2)
cross_val_model_w2G8_v2
```

## References


Adams, R. J., Wilson, M., & Wu, M. (1997). Multilevel item response models: An approach to errors in variables regression. *Journal of educational and behavioral Statistics, 22*(1), 47-76.

Aida, Y. (1994). Examination of Horwitz, Horwitz, and Cope's construct of foreign language anxiety: The case of students of Japanese. *The modern language journal, 78*(2), 155-168.

Ameringer, V. (2018). Cognitive Abilities: Different Memory Functions and Language Aptitude. In *Exploring Language Aptitude: Views from Psychology, the Language Sciences, and Cognitive Neuroscience* (pp. 19-42). Springer, Cham.

Ameringer, V., Green, L., Leisser, D., & Turker, S. (2018). Introduction: towards an interdisciplinary understanding of language aptitude. In *Exploring language aptitude: Views from psychology, the language sciences, and cognitive neuroscience* (pp. 1-15). Springer, Cham.

An, X., Hannum, E., & Sargent, T. (2007). Teaching quality and student outcomes: Academic achievement and educational engagement in rural Northwest China. *China: An International Journal, 5*(2), 309-334.

Anderson, R. C. (1984). Some reflections on the acquisition of knowledge. *Educational Researcher, 73*(10), 5-10. https://doi.org/10.3102/0013189X013009005.

Anderson, K. J., & Minke, K. M. (2007). Parent involvement in education: Toward an understanding of parents' decision making. *The Journal of Educational Research, 100*(5), 311-323. https://doi.org/10.3200/JOER.100.5.311-323.

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological review, 105*(1), 158.

Baker, F. B., & Kim, S. H. (Eds.). (2004). *Item response theory: Parameter estimation techniques*. CRC Press.

Baker, R. S., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *JEDM| Journal of Educational Data Mining, 1*(1), 3-17. https://doi.org/10.5281/zenodo.3554657.

Baker, S., & Inventado, P. S. (2016). Educational data mining and learning analytics: Potentials and possibilities for online education. In G. Veletsianos (Ed.), *Emergence and Innovation in Digital Learning* (83–98). doi:10.15215/aupress/9781771991490.01.

Bakhtin, M. M. (1981). *The dialogic imagination*. (C. Emerson & M. Holquist, trans.). Austin: University of Texas Press.

Bates, E., & MacWhinney, B. (1981). Second-language acquisition from a functionalist perspective: Pragmatic, semantic, and perceptual strategies. *Annals of the New York Academy of Sciences, 379*, 190–214. https://doi-org.oca.ucsc.edu/10.1111/j.1749-6632.1981.tb42009.x.

Beck, J. E., & Woolf, B. P. (2000, June). High-level student modeling with machine learning. In *International Conference on Intelligent Tutoring Systems* (pp. 584-593). Springer, Berlin, Heidelberg.

Berliner, D. C. (1988, February). *The development of expertise in pedagogy*. Charles W. Hunt Memorial Lecture presented at the Annual Meeting of the American Association of Colleges for Teacher Education (New Orleans, LA, February 17-20, 1988) ED 298 122.

Borko, H. and Livingston, C. (1989). Cognition and Improvisation: Differences in Mathematics Instruction by Expert and Novice Teachers. *American Educational Research Journal, 26*(4), 473-498. https://doi.org/10.3102/00028312026004473.

Buchmann, C., & Hannum, E. (2001). Education and stratification in developing countries: A review of theories and research. *Annual review of sociology, 27*(1), 77-102. https://doi.org/10.1146/annurev.soc.27.1.77.

Burden, P. R.(1982). Developmental supervision: Reducing teacher stress at different career stages. *U.S Department Of Education Educational Resources Information Center* [Paper presentation]. The 62th Association of Teacher Educators National Conference, Phoenix, AZ.

Butler, Y. G. (2014). Parental factors and early English education as a foreign language: A case study in Mainland China. *Research papers in education, 29*(4), 410-437.

Caldas, S. J., & Bankston, C. (1997). Effect of school population socioeconomic status on individual academic achievement. *The Journal of Educational Research, 90*(5), 269-277.

Cancio, E. J., West, R. P., & Young, K. R. (2004). Improving mathematics homework completion and accuracy of students with EBD through self-management and parent participation. *Journal of Emotional and Behavioral Disorders, 12*(1), 9-22. https://doi-org.oca.ucsc.edu/10.1177/10634266040120010201.

Carroll, J. B. (1958). A factor analysis of two foreign language aptitude batteries. *The Journal of General Psychology, 59*(1), 3–19. Doi: 10.1080/00221309.1958.9710168.

Carroll, J. B. (1964). The prediction of success in intensive foreign language training. In R. Glaser (Ed.), *Training research and education* (pp. 87–136). Pittsburgh, PA: University of Pittsburgh Press.

Carroll, J. B. (1967). Foreign language proficiency levels attained by language majors near graduation from college. *Foreign Language Annals, 1*(2), 131-151. https://doi-org.oca.ucsc.edu/10.1111/j.1944-9720.1967.tb00127.x.

Carroll, J. B., & Sapon, S. M. (1959). *Modern language aptitude test.* Psychological Corporation.

Carter, K., Cushing, K., Sabers, D., Stein, P., & Berliner, D. (1988). Expert-novice differences in perceiving and processing visual classroom stimuli. *Journal of Teacher Education, 39*(3), 25-31. https://doi.org/10.1177/002248718803900306.

Carter, K., Sabers, D., Cushing, K., Pinnegar, S., & Berliner, D. (1987). Processing and using information about students: A study of expert, novice and postulant teachers. *Teaching and Teacher Education, 3*(2), 147-157. https://doi.org/10.1016/0742-051X(87)90015-1.

Chen, Y. (2012). *Analysis and research of the ongoing status and trend of the established projects in Chinese education science development since the 21st century*. Shanghai: East China Normal University.

Chi, M., Feltovich, P, & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*(2), 121-152. https://doi.org/10.1207/s15516709cog0502_2.

Chihara, T., & Oiler Jr, J. W. (1978). Attitudes and attained proficiency in EFL: A sociolinguistic study of adult Japanese speakers. *Language learning, 28*(1), 55-68. https://doi-org.oca.ucsc.edu/10.1111/j.1467-1770.1978.tb00304.x.

Chomsky, N. (1965). *Aspects of a theory of syntax*. Cambridge, Mass.: MIT Press.

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.

Chomsky, N. (1986). *Knowledge of language*. New York: Praeger.

Christensen, J., Burke, P., Fessler, R., & Hagstrom, D. (1983). *Stages of Teachers' Careers: Implications for Professional Development*. Washington, DC: ERIC Clearinghouse on Teacher Education.

Christenson, S. L., Rounds, T., & Gorney, D. (1992). Family factors and student achievement: An avenue to increase students' success. *School Psychology Quarterly, 7*(3), 178.

Cole, M., John-Steiner, V., Scribner, S., & Souberman, E. (Eds.). (1978). *Mind in society: The development of higher psychological processes*. L. S. Vygotsky. Harvard U Press.

Cole, N. S. (1997). The ETS gender study: How females and males perform in educational settings.

Coleman, J. S., Campbell, E., Mood, A., Weinfeld, E., Hobson, C, York, R., & McPartland, J. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Government Printing Office.

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. Irvington.

Cummins, J. (1979). Cognitive/Academic Language Proficiency, Linguistic Interdependence, the Optimum Age Question and Some Other Matters. *Working Papers on Bilingualism*, No. 19.

Cummins, J. (1999). *BICS and CALP: Clarifying the Distinction*.

Darling-Hammond, L. (2002). The research and rhetoric on teacher certification: A response to “Teacher certification reconsidered.” *Educational Policy Analysis Archives, 10*(36), 1–55. https://doi.org/10.14507/epaa.v10n36.2002.

DeKeyser, R. M., & Koeth, J. (2011). Cognitive aptitudes for second language learning. *Handbook of research in second language teaching and learning, 2*, 395-406.

Dockrell, W. B., & Brosseau, J. F. (1967). Correlates of second language learning by young children. *Alberta Journal of Educational Research, 13*(4), 295-298.

Dörnyei, Z. (1990). Conceptualizing motivation in foreign-language learning. *Language learning, 40*(1), 45-78.

Dörnyei, Z. (2006). Individual differences in second language acquisition. *AILA Review, 19*(1), 42–68. https://doi.org/10.1075/aila.19.05dor.

Dossett, D., & Munoz, M. A. (2003). *Classroom accountability: A value-added methodology* [Paper presentation]. The annual meeting of the American Educational Research Association, Chicago, IL.

Duan, W., Guan, Y., & Bu, H. (2018). The effect of parental involvement and socioeconomic status on junior school students' academic achievement and school behavior in China. *Frontiers in psychology, 9*, 952.

Duckworth, A. L., & Seligman, M. E. (2006). Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores. *Journal of educational psychology, 98*(1), 198.

Dulay, H., & Burt, M. (1977). Remarks on Creativity in Language Acquisition. In M. Burt, H. Dulay, & M. Finocchiaro (Eds.), *Viewpoints on English as a Second Language* (pp. 95-126). New York: Regents.

Dulay, H., Burt, M.. & Krashen, S. (1982). *Language Two*. Oxford University Press, New York.

Ellis, N. (1996). Sequencing in SLA: Phonological Memory, Chunking, and Points of Order. *Studies in Second Language Acquisition, 18*(1), 91-126. doi:10.1017/S0272263100014698.

Ellis, N. (1999). Cognitive approaches to SLA. *Annual Review of Applied Linguistics, 19*, 22-42. https://doi-org.oca.ucsc.edu/10.1017/S0267190599190020.

Ellis, R. (1990). *Instructed second language acquisition*. Oxford: Basil Blackwell.

Elman, J. L., Bates, E. A., & Johnson, M. H. (1998). *Rethinking innateness: A connectionist perspective on development* (Vol. 10). MIT press.

Feiman-Nemser, S. (2012). *Teachers as learners*. Cambridge, MA: Harvard Education Press.

Fogarty, J., Wang, M. and Creek, R. (1983). A Descriptive Study of Experienced and Novice Teachers' Interactive Instructional Thoughts and Actions. *The Journal of Educational Research, 77*(1), 22-32. https://doi.org/10.1080/00220671.1983.10885491.

Fuchs, C., & Diamantopoulos, A. (2009). Using single-item measures for construct measurement in management research: Conceptual issues and application guidelines. *Die Betriebswirtschaft, 69*(2), 195.

Fuller, F. (1969). Concerns of teachers: A developmental conceptualization. American *Educational Research Journal, 6*(2), 207–226. https://doi.org/10.3102/00028312006002207.

Fuller, F., Parsons, J. S., & Watkins, J. E. (1973, December). *Concerns of teachers: Research and reconceptualization* [Paper presentation]. The 59th annual meeting of the American Educational Research Association. Chicago, Il.

Gardner, D. G., Cummings, L. L., Dunham, R. B., & Pierce, J. L. (1998). Single-item versus multiple-item measurement scales: An empirical comparison. *Educational and psychological measurement, 58*(6), 898-915.

Gardner, R. C., & Lambert, W. E. (1972). *Attitudes and Motivation in Second-Language Learning*. Rowley, MA: Newbury House.

Genesee, F. (1976). The role of intelligence in second language learning. *Language Learning, 26*(2), 267-280. https://doi.org/10.1111/j.1467-1770.1976.tb00277.x.

Greenberg, E., Rhodes, D., Ye, X. & Stancavage, F. (2004). *Prepared to teach: Teacher preparation and student achievement in eighth-grade mathematics*. Washington, DC: American Institute for Research.

Greenwald, R., Hedges, L. V., & Laine, R. D. (1996). The effect of school resources on student achievement. *Review of educational research, 66*(3), 361-396. https://doi.org/10.3102/00346543066003361.

Greenwood, G. E., & Hickman, C. W. (1991). Research and practice in parent involvement: Implications for teacher education. *The elementary school journal, 91*(3), 279-288.

Guo, Y., Chen, Q., Zhai, S., & Pei, C. (2020). Does private tutoring improve student learning in China? Evidence from the China Education Panel Survey. *Asia & the Pacific Policy Studies, 7*(3), 322-343.

Gutman, L. M., & Midgley, C. (2000). The role of protective factors in supporting the academic achievement of poor African American students during the middle school transition. *Journal of Youth and Adolescence, 29*(2), 233–248.

Halliday, M. A. K. (1975). Learning how to mean. In *Foundations of language development* (pp. 239-265). Academic Press.

Halliday, M. A. K. (1978). *Language as social semiotic: The social interpretation of language and meaning*. Hodder Arnold.

Halliday, M. A. (1993). Towards a language-based theory of learning. *Linguistics and Education*, 5(2), 93-116. https://doi-org.oca.ucsc.edu/10.1016/0898-5898(93)90026-7.

Harrison, G., & Kroll, L. (2007). Relationship between L1 and L2 word-level reading and phonological processing in adults learning English as a second language. *Journal of Research in Reading, 30*, 379-393. http://dx.doi.org/10.1111/j.1467-9817.2007.00351.x

Henderson, A. (1987). *The evidence continues to grow: Parent involvement improves student achievement*. Columbia, MD: National Committee for Citizens in Education.

Hill, N. E., & Tyson, D. F. (2009). Parental involvement in middle school: a meta-analytic assessment of the strategies that promote achievement. *Developmental psychology, 45*(3), 740.

Hu, G. (2002a). Recent important developments in secondary English-language teaching in the People's Republic of China. *Language, Culture and Curriculum, 15*, 30–49.

Hu, G. (2005). English language education in China: Policies, progress, and problems. *Language policy, 4*(1), 5-24.

Huang, F. L., & Moon, T. R. (2009). Is experience the best teacher? A multilevel analysis of teacher characteristics and student achievement in low performing schools. *Educational Assessment, Evaluation and Accountability, 21*(3), 209-234. https://doi-org.oca.ucsc.edu/10.1007/s11092-009-9074-2.

Hughes, J., & Kwok, O. M. (2007). Influence of student-teacher and parent-teacher relationships on lower achieving readers' engagement and achievement in the primary grades. *Journal of educational psychology, 99*(1), 39.

Jarvis, S., & Pavlenko, A. (2008). *Crosslinguistic influence in language and cognition*. Routledge.

Jessner, U. (2006). *Linguistic awareness in multilinguals: English as a third language*. Edinburgh, UK: Edinburgh University Press.

Jeynes, W. (2012). A meta-analysis of the efficacy of different types of parental involvement programs for urban students. *Urban education, 47*(4), 706-742.

Katz, L. (1972). Developmental Stages of Preschool Teachers. *The Elementary School Journal, 73*(1), 50-54. https://doi-org.oca.ucsc.edu/10.1086/460731.

Krashen, S. D. (1976). Formal and informal linguistic environments in language acquisition and language learning. *Tesol Quarterly 10*(2), 157-168. https://www.jstor.org/stable/3585637.

Krashen, S. D. (1982). *Principles and Practice in Second Language Acquisition*. Oxford: Pergamon.

Krashen, S., & Seliger, H. (1975). The essential contributions of formal instruction. *TESOL Quarterly 9*(2): 173-183. https://www.jstor.org/stable/3585484.

Krashen, S. D., Seliger, H., & Hartnett, D. (1974). Two studies in adult second language learning. *Kritikon Litterarum 2/3*, 220-228.

Ladd, H. F., & Sorensen, L. C. (2017). Returns to teacher experience: Student achievement and motivation in middle school. *Education Finance and Policy, 12*(2), 241-279. Doi: 10.1162/EDFP_a_00194.

Lai, F. (2010). Are boys left behind? The evolution of the gender achievement gap in Beijing's middle schools. *Economics of Education Review, 29*(3), 383-399.

Langacker, R. W. (1987). *Foundations of cognitive grammar: Theoretical prerequisites* (Vol. 1). Stanford university press.

Langacker, R. W. (1991). *Foundations of cognitive grammar: Descriptive application* (Vol. 2). Stanford university press.

Larsen-Freeman, D. I. A. N. E. (1991). Second language acquisition research: Staking out the territory. *TESOL quarterly, 25*(2), 315-350. https://www.jstor.org/stable/3587466.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge university press.

Leinhardt, G. (1986a). Expertise in mathematics teaching. *Educational Leadership, 43*(7), 28-33. Doi: 10.1007/978-1-4419-7707-6_3.

Leinhardt, G. & Greeno, J. (1986). The cognitive skill of teaching. *Journal of Educational Psychology, 78*(2), 75-95. https://doi.org/10.1037/0022-0663.78.2.75.

Li, M., Kirby, J. R., Cheng, L., Wade-Woolley, L., & Qiang, H. (2012). Cognitive predictors of English reading achievement in Chinese English-immersion students. *Reading Psychology, 33*(5), 423-447.

Li, Y., Hu, T., Ge, T., & Auden, E. (2019). The relationship between home-based parental involvement, parental educational expectation and academic performance of middle school students in mainland China: A mediation analysis of cognitive ability. *International Journal of Educational Research, 97*, 139-153.

Liu, J. (2021). Cognitive Returns to Having Better Educated Teachers: Evidence from the China Education Panel Survey. *Journal of Intelligence, 9*(4), 60.

Liu, R. (2018). Gender-math stereotype, biased self-assessment, and aspiration in STEM careers: The gender gap among early adolescents in China. *Comparative Education Review, 62*(4), 522-541.

Livingston, C., & Borko, H. (1990). High School Mathematics Review Lessons: Expert-Novice Distinctions. *Journal for Research in Mathematics Education, 21*(5), 372. DOI: 10.2307/749395.

Lortie, D. C. (1966). Teacher socialization: The Robinson Crusoe model. In *The real world of the beginning teacher*. Washington, DC. National Commission on Teacher Education and Professional Standards (ERIC Document Reproduction Service No. ED 030 616).

Luo, R., Teng, K., Xu, L., & Zhu, Y. (2021). A decomposition of student achievement gap by gender in China: Evidence from random class assignment. *International Journal of Educational Research, 106*, 101721.

Maccoby, E. E., & Martin, J. A. (1983). Socialization in the context of the family: parent-child interaction. In P. H. Mussen (ed.) & E. M. Hetherington (Vol. ed.), *Handbook of child psychology, Vol. 4,* (pp. 1-101). New York: Wiley.

Maluch, J. T., Kempert, S., Neumann, M., & Stanat, P. (2015). The effect of speaking a minority language at home on foreign language learning. *Learning and instruction, 36*, 76-85.

Mei, Y., & Cheng, Q. (1954). A Survey of Chinese Students in American Universities and Colleges in the Past One Hundred Years, under the Joint Sponsorship of National Tsing Hua University Research Fellowship Fund and China Institute in America, New York.

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior.* Henry Holt and Co. https://doi.org/10.1037/10039-000.

Minke, K. M., Sheridan, S. M., Kim, E. M., Ryoo, J. H., & Koziol, N. A. (2014). Congruence in parent-teacher relationships: The role of shared perceptions. *The elementary school journal, 114*(4), 527-546.

Mosteller, F., & Moynihan, D. P. (1972). *On equality of educational opportunity*. New York: Random House.

National Bureau of Statistics of China. (2019). *21. Education.* Retrieved from http://www.stats.gov.cn/tjsj/ndsj/2019/indexch.htm.

Ochs, E. (1988). *Culture and language development*. New York: Cambridge University Press.

OECD. (2016). *Education in China: A snapshot*. Retrieved from https://www.oecd.org/china/Education-in-China-a-snapshot.pdf.

Onwuegbuzie, A. J., Bailey, P., & Daley, C. E. (2000). Cognitive, affective, personality, and demographic predictors of foreign-language achievement. *The Journal of Educational Research, 94*(1), 3-15.

Ortega, L. (2014). *Understanding second language acquisition*. Routledge.

Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford, UK: Oxford University Press.

Phillipson, S., & Phillipson, S. N. (2007). Academic expectations, belief of ability, and involvement by parents as predictors of child achievement: A cross-cultural comparison. *Educational Psychology, 27*(3), 329-348.

Proctor, C. P., August, D., Carlo, M., & Snow, C. (2006). The intriguing role of Spanish language vocabulary knowledge in predicting English reading comprehension. *Journal of Educational Psychology, 98*(1), 159-69. http://dx.doi.org/10.1037/0022-0663.98.1.159.

Rice, J. K. (2003). *Teacher quality: Understanding the effectiveness of teacher attributes*. Washington, DC: Economic Policy Institute.

Romero, C., & Ventura, S. (2010). Educational data mining: a review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 40*(6), 601-618. 10.1109/TSMCC.2010.2053532.

Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of research and development,* 44, 206-226. Doi: 10.1147/rd.441.0206.

Sears, R. R., Maccoby, E. E., & Levin, H. (1957). *Patterns of child-rearing*. Evanston, IL: Row, Peterson.

Shavelson, R. J. (1986, June). *Interactive decision making: Some thoughts on teacher cognition*. Invited address, I Congreso

International, "Pensamientos de los Profesores y Toma de Decisiones," Seville, Spain.

Shaver, A. V., & Walls, R. T. (1998). Effect of Title I parent involvement on student reading and mathematics achievement. *Journal of Research & Development in Education, 31*(2), 90-97.

Sheldon, S. B. (2003). Linking school–family–community partnerships in urban elementary schools to student achievement on state tests. *The Urban Review, 35*(2), 149-165. 10.1023/A:1023713829693.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*(1), 1-22. https://doi.org/10.17763/haer.57.1.j463w79r56455411.

Shun, K. L., & Wong, D. B. (Eds.). (2004). *Confucian ethics: A comparative study of self, autonomy, and community.* Cambridge University Press.

Singleton, D. (2017). Language aptitude: Desirable trait or acquirable attribute? *Studies in Second Language Learning and Teaching, 7*(1), 89-103. https://doi.org/10.14746/ssllt.2017.7.1.5.

Slobin, D. I. (1997). The origins of grammaticizable notions: Beyond the individual mind. In D. I. Slobin (ed.) *The crosslinguistic study of language acquisition, Vol. 5*. Mahway, NJ: Erlbaum. 265-323.

Spera, C. (2005). A review of the relationship among parenting practices, parenting styles, and adolescent school achievement. *Educational psychology review, 17*(2), 125-146.

Sun, Y. (2020). Migration and education inequality in China: Evidence from China Education Panel Study. *International Journal of Education Humanities and Social Science, 3*(4), 350-366.

Tang, T. Y., & McCalla, G. (2003, July). Smart recommendation for an evolving e-learning system. In *Workshop on Technologies for*

*Electronic Documents for Supporting Learning, International Conference on Artificial Intelligence in Education* (pp. 699-710).

Tani, M., Xu, L., & Zhu, Y. (2021). The impact of an un (der) funded inclusive education policy: Evidence from the 2013 China education panel survey. *Journal of Economic Behavior & Organization, 190*, 768-784.

Thøgersen, S. (2002). *A County of Culture: Twentieth-century China Seen from the Village Schools of Zouping, Shangdong.* Ann Arbor, The University of Michigan Press.

Turner, E. A., Chandler, M., & Heffer, R. W. (2009). The influence of parenting styles, achievement motivation, and self-efficacy on academic performance in college students. *Journal of college student development, 50*(3), 337-346.

Wang, F. (2008). Motivation and English achievement: An exploratory and confirmatory factor analysis of a new measure for Chinese students of English learning. *North American journal of psychology, 10*(3), 633-646.

Wang, T. (2021). Classroom Composition and Student Academic Achievement: The Impact of Peers' Parental Education. *The BE Journal of Economic Analysis & Policy, 21*(1), 273-305.

Wang, W., & Gao, X. (2008). English language education in China: A review of selected research. *Journal of Multilingual and Multicultural Development, 29*(5), 380-399.

Watson-Gegeo, K. A. (2004). Mind, language, and epistemology: Toward a language socialization paradigm for SLA. *The Modern Language Journal, 88*(3), 331-350. https://doi-org.oca.ucsc.edu/10.1111/j.0026-7902.2004.00233.x.

Wayne, A. J. & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research, 73*(1), 89–122. https://doi-org.oca.ucsc.edu/10.3102/00346543073001089.

Wen, Q., & Johnson, R. K. (1997). L2 learner variables and English achievement: A study of tertiary-level English majors in China. *Applied linguistics, 18*(1), 27-48.

Wen, W. (2005). A review of the development of higher education studies in China. *Higher Educational Studies, 26*(6), 54-59.

Wen, W., & Xie, W. (2017). The development and characteristics of educational studies in China. *Knowledge and the study of education: An international exploration*, 145-160.

Westerman, D. (1991). Expert and Novice Teacher Decision Making. *Journal of Teacher Education, 42*(4), 292-305. https://doi-org.oca.ucsc.edu/10.1177/002248719104200407.

White, K. R. (1982). The relationship between socioeconomic status and academic achievement. *Psychological Bulletin, 91*(3) , 461–481.

Wong, D. (2020). Comparative philosophy: Chinese and western. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*. Retrieved from: https://plato.stanford.edu/archives/fall2020/entries/comparphil-chiwes/.

Xia, Y. (2019). *Perceptions on Language and Literacy Integration in Mathematics Instruction: A Study of Expert and Novice Teacher*. Unpublished manuscript, University of California, Santa Cruz. CA.

Xu, D., & Li, Q. (2018). Gender achievement gaps among Chinese middle school students and the role of teachers' gender. *Economics of Education Review, 67*, 82-93.

Yang, J., & Zhao, X. (2021). Does all work and no play make elite students? Evidence from the China education panel survey. *International Journal of Educational Development, 80*, 102321.

Yang, R. Internationalisation, Indigenisation and Educational Research in China. *Australian Journal of Education, 49*(1), 66-88. https://doi.org/10.1177/000494410504900104.

Ye, L. (2004). Development of pedagogy in China: A review of century-long issues. *Jiaoyu Yanjiu [Educational Research], 7*, 3-17.

Yinger, R. J. (1987, April). *By the seat of your pants: An inquiry into improvisation and teaching* [Paper presentation]. The 71th annual meeting of the American Educational Research Association, Washington, DC.

Young, W., Weckman, G., & Holland, W. (2011). A survey of methodologies for the treatment of missing values within datasets: Limitations and benefits. *Theoretical Issues in Ergonomics Science, 12*, 15–43. Doi:10.1080/14639220903470205.

Yue, C., & Xu, X. (2019). Review of Quantitative Methods Used in Chinese Educational Research, 1978–2018. *ECNU Review of Education, 2*(4), 515-543.

Yuxiao, W., & Chao, H. (2017). School socioeconomic segregation and educational expectations of students in China's junior high schools. *Social Sciences in China, 38*(3), 112-126.

Zaïane, O. (2002). Building A Recommender Agent for e-Learning Systems. In *Proceedings of the International Conference in Education* (pp. 55-59). Auckland, New Zealand.

Zaïane, o., Luo, J. (2001). Web usage mining for a better web-based learning environment. In *Proceedings of Conference on Advanced Technology for Education* (pp. 60-64). Banff, Alberta.

Zhao, Y., Zhang, G., Yang, W., Kirkland, D., Han, X., & Zhang, J. (2008). A comparative study of educational research in China and the United States. *Asia Pacific Journal of Education, 28*(1), 1-17. https://doi.org/10.1080/02188790701849826.

Zheng, R., & Cui, L. (2001). A review of educational research methods in the last two decades. *Jiaoyu Yanjiu [Educational Research]*, 6, 17–21.

Zhu, Y. (Ed.). (1988). Documents on educational system in modern China (2nd ed.). Huazhong: Huazhong Normal University Press.

Zuengler, J., & Miller, E. R. (2006). Cognitive and sociocultural perspectives: Two parallel SLA worlds?. *TESOL quarterly, 40*(1), 35-58. https://www.jstor.org/stable/40264510.