# Factors Associated with Individual Differences in Reading Performance of 10 Year Olds

Olivia Jeffris

Committee: Dr. Terry Jernigan, Dr. Douglas Nitz, and Dr. Clare Palmer

Cognitive Science Honors Program 2020-2021

## Abstract

This project aims to generate a model that explains individual variability in reading performance among participants of the ABCD study at baseline. The ABCD Study is a national longitudinal study that collects genetic, neuroimaging, physiological, and behavioral data from 11,880 children aged 9-10 years to track factors that change over adolescent development. The three main aims of this project are: (1) to determine the extent to which variables across socioeconomic status (SES), individual traits, and environmental traits can explain individual variability in reading; (2) to determine the extent to which variability in reading performance is shared amongst these predictors; and (3) to determine the extent to which SES mediates the associations between these predictors and reading. In order to address these aims, three different sets of linear regression models were run. The first set of models were run to determine the amount of variability in reading score explained by three different models: one null model in which age and sex predicted reading score; another null model in which age, sex, and SES predicted reading score; and a full model in which all variables of interest predicted reading score. The variables of interest all fall into the following behavioral domains: Linguistic Development Factors, Bilinguality Factors, Recreational Reading, Academic Factors, Screen Time Factors, Sleep Factors, Psychopathology/Temperament Factors, School/Neighborhood Environment Factors, Psychosocial Environment Factors, ABCD Site, SES, and Age and Sex. The second set of models were run to determine the variance in reading score explained by individual variables in these domains, controlling for age and sex. The third set of models were run to determine the variability in reading score explained by each behavioral domain and the extent to which these domains overlap with SES and each other in the variability they account for in reading score. After running these analyses, Recreational Reading, Academic Factors, and SES yielded the largest effect sizes on their own, meaning they explained the most variance in reading score when controlling for all other factors. All of the behavioral domains were mediated by SES and each other to some degree, suggesting that variability in reading performance is shared amongst these domains.

# Introduction

In spite of much emphasis in early childhood education on literacy skills and reading instruction, there is lots of variability observed in the reading skills of 9-10 year old children. Numerous studies have been conducted on early literacy, but there remains little consensus about the most important factors contributing to this variability. Better models of the factors that may influence developing reading skills are needed to identify the best targets for interventions that would improve literacy in a diverse population of adolescents. With this in mind, my research questions are: (1) How much variability in reading performance can be predicted from variables in the ABCD Study? and (2) Does the variability in reading performance predicted by these variables overlap with SES?

Many studies have been conducted to investigate factors associated with school-aged reading performance. A few of such studies are highlighted in the following paragraphs.

Socioeconomic status (SES), defined as family income and highest level of parent education, is associated with school-aged reading performance. Psychosocial literature has largely declared that family income is a predictor of adolescent reading abilities, specifying that poverty negatively impacts adolescent reading outcomes (Bhattacharya, 2010). Sociocultural literature suggests that impoverished children have poorer literacy because they don't receive sufficient literacy experiences in their homes (Bhattacharya, 2010). Van Zwieten et al. further confirmed that low SES was associated with lower reading performance in school-aged children (2020). Overall, there has been a large consensus on the existence of a positive association between adolescent SES and reading performance.

Studies show that a child's physical environment, such as their neighborhood, also relates to their reading performance. A study conducted by Froiland et al. determined that neighborhood SES predicted home literacy, defined as parents and children reading together and the number of books in a home, which subsequently predicted early literacy abilities in children (2013b). Another study found that children who lived in affluent neighborhoods during early childhood exhibited higher academic achievement (Anderson et al, 2014). This study also concluded that longer periods of time spent living in affluent neighborhoods correlated with children's reading and math achievement later in adolescence (Anderson et al, 2014). Overall, findings have suggested that the extent to which children live in affluent or impoverished neighborhoods is associated with their reading performance.

Studies have also investigated the relationship between linguistic development and school-aged reading performance. Most studies conducted on this topic agree that the age of speech development is related to a child's subsequent reading performance. Rescorla found that delayed speech onset in participants was associated with lower performance on vocabulary, grammar, verbal memory, and reading comprehension measures administered to the participants at age 13 (2005). Furthermore, Preston reported that children who spoke earlier than average scored highest on standardized measures of language, reading, and spelling while children who spoke later than average scored lowest on these measures (Preston et al, 2010). Studies

conducted on this topic agree that there is an inverse correlation between the early linguistic development of a child and their later reading performance.

Bilinguality has been identified as a correlate of school-aged reading performance as well. For bilingual children in the American education system, developing English language skills is crucial to their literary achievement in school. In one study, Spanish/English bilingual kindergarten students were assessed using standardized vocabulary and narrative elicitation tests and, despite improving their English vocabulary performance from kindergarten to first grade, most of these students still scored lower than their English monolingual peers in first grade (Uccelli and Paez, 2007). The bilingual children in this study struggled to develop English literacy skills comparable to their monolingual counterparts when receiving instruction in English. As such, some studies have suggested that bilingual and English instruction may have varying effects on the reading proficiency of bilingual children. Another study showed that bilingual children receiving English instruction had better Spanish and English language proficiency scores than those receiving bilingual instruction (Lindholm-Leary, 2012). Additionally, bilingual children who entered the program as Mostly Proficient in Spanish had higher English language proficiency scores than those who entered as Mostly Limited in Spanish (Lindholm-Leary, 2012). The intersection between adolescent bilinguality and the language in which they receive instruction has varying associations with their reading proficiency.

The extent to which school-aged children read for recreational purposes has also been shown to correlate with their reading performance. Froiland found that a child's intrinsic motivation to read is associated with their reading performance (2013a). Another study shows that reading for pleasure is significantly associated with cognitive progress in vocabulary and mathematics measures between ages 10 and 16 (Sullivan and Brown, 2015). Recreational reading has been established as a positive correlate of adolescent reading performance.

Academic factors, such as enrollment in gifted programs in school, have also been observed as potential correlates of school-aged reading performance. Rowe et al. reported that the test scores of adolescent participants enrolled in gifted and talented school programs on the Full Scale IQ (FSIQ), General Ability Index (GAI), Verbal Comprehension exam, and Perceptual Reasoning exam were associated with their reading comprehension standardized test scores (2012). However, there is lots of controversy over the predictive power of gifted program enrollment on reading performance in school-aged children because the students enrolled in these programs have been shown to possess heterogeneous reading abilities in some cases. One study showed that close to 10% of students in gifted programs read two or more years below their grade level and at 1 or more standard deviations below the mean for their age in at least one of the following domains: reading prose accuracy, reading prose comprehension, or isolated word reading accuracy (Munro, 2009). Additionally, a case study on children in gifted programs with learning disabilities discovered 3 'profiles' of these children: students gifted verbally, nonverbally, and in both areas (Munro, 2005). These findings suggest that students in gifted programs vary drastically in their literacy strengths, weaknesses, and abilities and that, as a

result, enrollment in gifted programs may not be a reliable indicator of school-aged reading performance.

Studies have also found associations between the time children spend engaged with screens, such as watching television and playing video games, and their reading habits. One study uncovered a negative association between the time children spend reading and the time they spend watching screens (Han, 2019). Other studies have identified similar relationships, such as a study conducted by Zebroff and Kaufman which concluded that youths spend more time on screen-based activities such as texting and less time on traditional reading (2017). Degree of screen time consumption has been established as a negative correlate of school-aged reading performance.

Prior studies have also identified childhood psychopathology and temperament as predictors of school-aged reading performance. One study found that adolescents with poor reading skills had higher levels of ADHD and anxiety disorders and more functional impairment across many areas when compared to adolescents in their cohort with average reading skills (Goldston et al, 2007). Many studies have suggested that temperament is associated with adolescent reading performance. Because anxiety is multifaceted, anxiety traits have mixed associations with reading performance. For example, one study looked at first grade participants and found that reading fluency was negatively correlated with separation anxiety symptoms while reading decoding was positively correlated with harm avoidance anxiety symptoms (Grills-Taquechel et al, 2012). They also found that harm avoidance and total anxiety symptoms, for females only, were positively correlated with reading fluency (Grills-Taquechel et al, 2012). Given the diverse and variable nature of psychopathology and temperament, these factors have been shown to vary in their relationships with school-aged reading performance.

After reviewing studies that highlight certain factors associated with school-aged reading performance, I reviewed studies that investigated relationships between these factors. Some of those studies are highlighted in the following paragraphs.

Researchers have observed a comorbidity of early linguistic problems and childhood psychopathology. Carpenter and Drabick built upon this finding by creating an all-encompassing model that combines three different models to explain the comorbidity of language and behavioral problems in children: (1) language difficulties influence risk for behavioral problems, (2) behavioral problems influence risk for language difficulties, and (3) other shared risk factors explain the comorbidity between behavior and language difficulties (2010). Some studies have found more specific links between ADHD and linguistic problems in children. Tennenbaum et al. observed that a cohort of children with varying developmental disorders possessed the following traits: inattention, hyperactivity, IQ development quotients above 70, and speech delays (1993). Inattention and hyperactivity are symptoms of ADHD and they are reported to co-occur with speech delay in some instances. Although the studies on this topic are limited, links have been established between linguistic development and child psychopathology.

Enrollment in gifted and talented programs has a relationship to some of the factors outlined earlier in this paper. Esquierdo and Arreguín-Anderson report that Hispanic bilingual

children are underrepresented in gifted programs (2012). Additionally, gifted children are reported to spend more time reading for pleasure than for information (Bobel, 1981). These findings highlight associations between gifted programs, bilinguality, and recreational reading that can be further investigated.

Socioeconomic status has variable relationships with the factors outlined earlier in this paper. A study conducted by Peverill et al. concluded that children with lower SES were more likely to develop psychopathology; that youth who endured intense financial hardship had the most psychopathology, and that SES was more strongly associated with behavioral problems than anxiety and depression (2021). Other studies have investigated the effects of SES on academic programs. For example, a study conducted by McBee found that children from poorer backgrounds are underrepresented in gifted programs (2006). Lastly, there are established links between SES, neighborhood environment, and reading performance in children. Willms suggests that, when looking at reading performance in children from varying SES backgrounds, socioeconomic gradients "converge at higher levels of socioeconomic status" (2003). As a result, youth from high SES backgrounds succeed in their reading performance regardless of where they live, while youth from low SES backgrounds have more variance in their reading performance (Willms, 2003). SES appears to be a correlate of childhood psychopathology, gifted program enrollment, and neighborhoods.

Overall, the factors outlined in this section are largely interrelated and have varying relationships to one another. Previous studies have looked at each of these factors as having an individual effect on reading performance, but the overlapping variability in reading performance explained by multiple factors remains largely uninvestigated.

The main aim of this project is to generate a model that explains individual variability in reading performance among participants of the ABCD study at baseline. The ABCD Study is a national longitudinal study that collects genetic, neuroimaging, physiological, and behavioral data from 11,880 children aged 9-10 years to track factors that change over adolescent development. Participants were recruited to reflect the national population as closely as possible. Therefore, participants are demographically diverse, allowing for an in-depth understanding of neurodevelopment across all sectors of society. The participants are assessed every year, allowing the ABCD Study research team to track variability among adolescents throughout the course of their development.

The main outcome measure used will be the participants' oral reading scores on the NIH Toolbox Oral Reading Recognition Test. The NIH Toolbox Oral Reading Recognition Test is administered by presenting letters and words to participants and having the participants read and pronounce them as accurately as possible. The predictors of interest fall under the following behavioral domains: Linguistic Development Factors, Bilinguality Factors, Recreational Reading, Academic Factors, Screen Time Factors, Sleep Factors, Psychopathology/Temperament Factors, School/Neighborhood Environment Factors, Psychosocial Environment Factors, and SES.

The three main aims of this project are: (1) to determine the extent to which variables across SES, individual traits, and environmental traits can explain individual variability in reading; (2) to determine the extent to which variability in reading performance is shared amongst these predictors; and (3) to determine the extent to which SES mediates the associations between these predictors and reading.

By conducting this data analysis project, I was able to pinpoint relationships that exist between oral reading scores and many other variables that describe the individual and environmental contexts of ABCD Study participants. My analysis project has the potential to identify attributes in children that account for variability in their reading abilities. These discoveries will illuminate some important factors that contribute to variability in adolescent reading performance and identify potential targets for interventions to improve adolescent literacy.

# Methods

*Variables*

In this study, a linear regression model was generated that explains individual variability in reading performance among participants of the ABCD study at baseline. All of the predictors of interest, covariates, and outcome variables used in this model have either been directly taken from the ABCD Study dataframe or have been transformed and adapted from variables in the dataframe. The main outcome measure used in the model is the NIH Toolbox Oral Reading Recognition Test Score. Many of the predictors of interest are summary scores extracted from questionnaires and scales such as the Child Behavior Checklist (CBCL) or stand-alone variables such as the parent-reported age of a youth participants' first word. The summary scores are generated by adding different subsets of item-level variables from each scale or questionnaire. The specified covariates in the model include age, birth-assigned sex, level of highest parent education, and total combined family income. The covariates are nuisance variables that are controlled for in order to reduce potential confounds in the model. Level of parent education and total combined family income serve as proxies for socioeconomic status (SES) in the model and they serve as variables of interest in some of the analyses run.

During analysis, the predictors of interest were either studied individually or as part of a behavioral domain. These domains were formed based on the similarity of the constructs the variables measure. The behavioral domains are as follows:

| **Domain** | **Instruments** | **Variables** |
|---|---|---|
| Linguistic Development Factors | ABCD Developmental History | First Word; Speech Delay |

| | | |
|---|---|---|
| Bilinguality Factors | PhexX Acculturation Survey | Parent English Proficiency; Youth English Proficiency; Parent Bilinguality; Youth Bilinguality |
| Recreational Reading | ABCD Sports Activities Read/Music - Parent | Recreational Reading |
| Academic Factors | ABCD Parent Diagnostic Interview for DSM-5 Background Items Full | KSADS Special Education; KSADS Gifted Program |
| Screen Time Factors | ABCD Youth Screen Time Survey | TV Screen Time; Videogames Screen Time; Video Screen Time; Texting Screen Time; Social Media Screen Time; Videochatting Screen Time |
| Sleep Factors | ABCD Parent Sleep Disturbance Scale for Children; Children's Report of Sleep Patterns | Average Sleep; Sleep Disturbances |
| Psychopathology/Temperament Factors | ABCD Parent Child Behavior Checklist Scores Aseba; ABCD Youth Wills Problem Solving Scale; Behavioral Inhibition Scale / Behavioral Activation Scale; UPPS Impulsive Behavior Scale | CBCL Aggressive Syndrome; CBCL Anxiety/Depression Syndrome; CBCL Rulebreak Syndrome; CBCL Attention Syndrome; CBCL Social Syndrome; CBCL Thought Syndrome; CBCL Somatic Syndrome CBCL Withdrawn/ Depressed Syndrome; Problem Solving; BAS Drive; BAS Fun Seeking; BAS Reward Responsiveness; BIS Summary Score; UPPS Lack of Perseverance; UPPS Lack of Planning; UPPS Positive Urgency; UPPS Negative Urgency; UPPS Sensation Seeking |

| | | |
|---|---|---|
| School/Neighborhood Environment | School Risk & Protective Factors Survey; ABCD Parent Neighborhood Safety / Crime Survey | Neighborhood Safety & Crime; School Disengagement; School Involvement; School Environment |
| Psychosocial Environment | Parental Monitoring Survey; PhenX Family Environment Scale - Family Conflict; Perceived Discrimination Scale; ABCD Children's Report of Parental Behavioral Inventory; Youth Prosocial Behavior Survey; Parent Prosocial Behavior Survey; ABCD Other Resilience | Parental Monitoring; Parent Family Conflict; Youth Family Conflict; Perceived Discrimination; Parental Acceptance; Youth Prosociality; Parent Prosociality; Friends; Close Friends |
| Socioeconomic Status (SES) | | Highest Level of Parent Education, Combined Family Income |
| ABCD Site | | ABCD Site |
| Age & sex | | Age, Sex |

A comprehensive list of the predictor, covariate, and outcome variables can be found in Appendix A with descriptions of what they measure and any transformations applied to them.

*Data Cleaning*

Data cleaning consisted of five steps. First, each variable of interest was extracted from the comprehensive ABCD Study database that contains all variables collected through the study. Some variables were initially composed of character values in the ABCD Study dataframe, such as "false" or "true" or "strongly disagree", "disagree", "neither agree nor disagree", "agree", and "strongly agree". So, numeric transformations were performed on these variables using the as.numeric() function in RStudio. All of the numeric transformations applied are outlined in the Variable Table of Appendix A. Performing these numeric transformations enabled these variables to be included in statistical analyses with other numeric variables.

Another important data cleaning step was setting thresholds for certain variables. I detected and pruned outliers in the data that were likely to be errors so that the analyses were not heavily influenced by outlier data. These thresholds were set by omitting highly unlikely or

impossible values based on what the variable measured. All of the thresholds set for each variable are outlined in the Variable Table of Appendix A.

Once all of the predictor, covariate, and outcome variables were extracted and properly transformed, all of the variables were combined into a single dataframe, denoted the "summary dataframe". The null values were omitted from the data frame to ensure that each column in the summary dataframe had the same number of values.

*Linear Regression Models*

Linear regression models were run using the gamm4 package in RStudio. For each of these models, a combination of covariates and predictors of interest served as the independent variables and reading score served as the dependent variable. Each linear regression model yields an $R^2$ value, which indicates the amount of variability that a given set of predictors can explain in an outcome measure.

In the first set of analyses, I aimed to analyze the variability in reading predicted by three different sets of variables. I ran three separate linear regression models:

$$\underline{\text{Model 1}}: \textit{reading score} \sim \textit{age} + \textit{sex} + (1|\textit{family}) + \varepsilon$$
$$\underline{\text{Model 2}}: \textit{reading score} \sim \textit{age} + \textit{sex} + \textit{SES} + (1|\textit{family}) + \varepsilon$$
$$\underline{\text{Model 3}}: \textit{reading score} \sim \textit{all predictor variables} + (1|\textit{family}) + \varepsilon$$

The 1|family term is the random effect that accounts for the relatedness between families. Including this in the model allows for a different intercept to be created for each family. The ε term is the error term, or the variance in the dependent variable not explained by the model.

In the first model, age and sex were predictors and reading score was the outcome variable. In the second model age, sex, and SES were predictors and reading score was the outcome variable. In the third model all of the predictors of interest and the covariates were predictors and reading score was the outcome variable. The $R^2$ values associated with each model were recorded. These $R^2$ values measure how much variance in reading score each of the three models explain overall.

In the second set of analyses, I aimed to analyze the variability in reading predicted by each variable, above and beyond age and sex. I ran 51 separate linear regression models. In each model age, sex, and one predictor were included as predictors and reading score was the outcome variable. This process was repeated 51 times to produce models with age, sex, and each of the remaining variables in the dataframe. The $R^2$ values associated with each model were recorded. Once all of the models had been run, I measured the change in $R^2$ from the model in which age and sex predict reading score to a model additionally including age, sex, and a predictor of interest for each of the 51 models. These $\Delta R^2$ values measure how much variability in reading score each predictor of interest accounts for on its own, controlling for age and sex. Age and sex are controlled for because they are nuisance variables.

After measuring the $\Delta R^2$ values for each of these variables, I ran likelihood ratio tests with a p value of 0.00096154. Running these tests allowed me to determine which $\Delta R^2$ values were statistically significant.

In the final set of analyses, I aimed to analyze the variability in reading predicted by sets of variables in behavioral domains. I ran 42 separate linear regression models. First, 14 linear regression models were run in which reading was predicted by age, sex, and one set of variables probing a similar behavioral domain at a time. [Refer to above Variables section for information on the behavioral domains]. The $R^2$ values associated with each model were recorded. Once all of the models had been run, I measured the change in $R^2$ from the model in which age and sex predict reading score to a model additionally including age, sex, and a set of variables in a behavioral domain for each of the 14 models. These $\Delta R^2$ values show how much variability in reading each domain of variables accounts for on their own, controlling for age and sex.

Then, 13 linear regression models were run in which reading was predicted by age, sex, SES, and one domain of variables at a time. The $R^2$ values associated with each model were recorded. After running all of the models, I measured the change in $R^2$ from the model in which age, sex, and SES predict reading score to a model additionally including age, sex, SES, and a set of variables in a behavioral domain for each of the 13 models. These $\Delta R^2$ values measure the unique $R^2$ associated with the variable of interest when controlling for age, sex, and SES.

Lastly, 15 linear regression models were run in which every domain of variables except one were included as predictors. This process was repeated, temporarily omitting a single domain of variables one at a time while all of the other predictors remained in the model. The $R^2$ values associated with each model were recorded. After running all of the models, I measured the change in $R^2$ from the model in which all of the variables predict reading score to a model in which a single set of variables in a behavioral domain is removed. These $\Delta R^2$ values measure the unique variability associated with the removed variable, controlling for all other predictors.

# Results

The first set of analyses yielded an overall $R^2$ value associated with three different linear regression models predicting reading score (Figure 1). The $R^2$ value associated with the model in which age and sex predict reading score was 4.84%. The $R^2$ value associated with the model in which age, sex, and SES predict reading score is 14.1%. The $R^2$ value in which all predictors of interest and covariates predict reading score is 33.8%. As more predictors were added to the model, the $R^2$ value associated with the model increased.

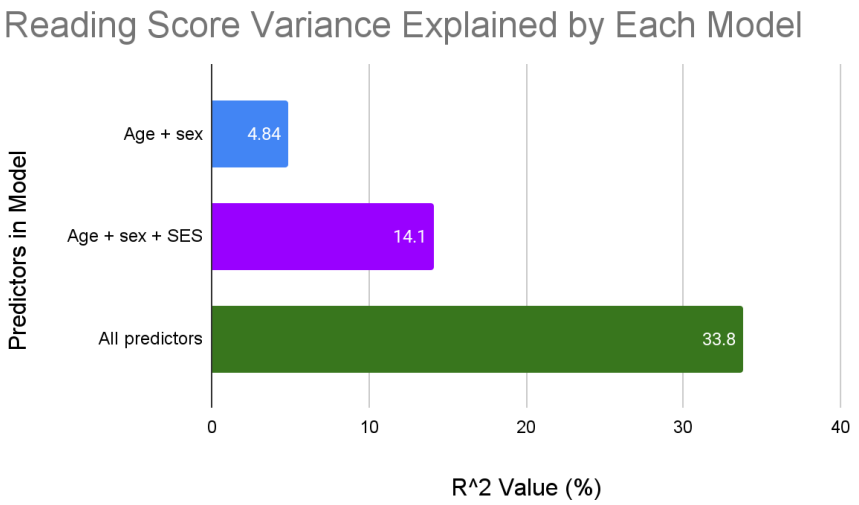


**Figure 1:** $R^2$ Value Associated with Each Model

The y-axis shows which predictors are included in each linear regression model. The x-axis represents the $R^2$ value, in percentages, associated with each model. In each model, reading score is the outcome measure. The numeric labels on the rightmost side of each bar represent the $R^2$ value associated with the given model. As more predictors are added to the model, the $R^2$ value increases.

The second set of analyses measured the $\Delta R^2$ from a linear regression model with age and sex predicting reading score to a linear regression model with age, sex, and one variable of interest predicting reading score (Figure 2). The largest effect sizes resulted from adding the following predictors to the model: Speech Delay; Recreational Reading; KSADS Special Education, KSADS Gifted Program; and SES. When Speech Delay is added, $R^2$ changes by 0.026. When Recreational Reading is added, the $R^2$ changes by 0.1027. When KSADS Special Education is added, the $R^2$ changes by 0.0807. Lastly, when KSADS Gifted Program is added, the $R^2$ changes by 0.0517. When SES is added, the $R^2$ changes by 0.0807.

Some smaller effect sizes that still yielded interesting findings resulted from adding the following variables: Average Sleep; TV Screen Time; Video Screen Time; Youth English Proficiency; Speech Delay; Problem Solving; CBCL Rulebreak Syndrome; CBCL Attention Syndrome; CBCL Social Syndrome; UPPS Positive Urgency; Parental Monitoring; and Perceived Discrimination. When Average Sleep is added, $R^2$ changes by 0.0181. When TV Screen Time is added, $R^2$ changes by 0.0204. When Video Screen Time is added, $R^2$ changes by 0.018. When Youth English Proficiency is added, $R^2$ changes by 0.0068. When Problem Solving is added, $R^2$ changes by 0.016. When CBCL Rulebreak Syndrome is added, $R^2$ changes by 0.0092. When CBCL Attention Syndrome is added, $R^2$ changes by 0.0124. When CBCL Social Syndrome is added, $R^2$ changes by 0.0094. When UPPS Positive Urgency is added, $R^2$ changes by 0.087. When Parental Monitoring is added, $R^2$ changes by 0.0085. Lastly, when Perceived Discrimination is added, $R^2$ changes by 0.014.

After running likelihood ratio tests, the all of the variables except for the following yielded statistically significant $\Delta R^2$ values: First Word, Sleep Disturbances, CBCL Anxiety/Depression Syndrome, CBCL Thought Syndrome, CBCL Somatic Syndrome, CBCL Withdrawn/Depressed Syndrome, BIS Summary Score, UPPS Lack of Planning, School Involvement, School Environment, Youth Prosociality, Parent Prosociality, and Friends.

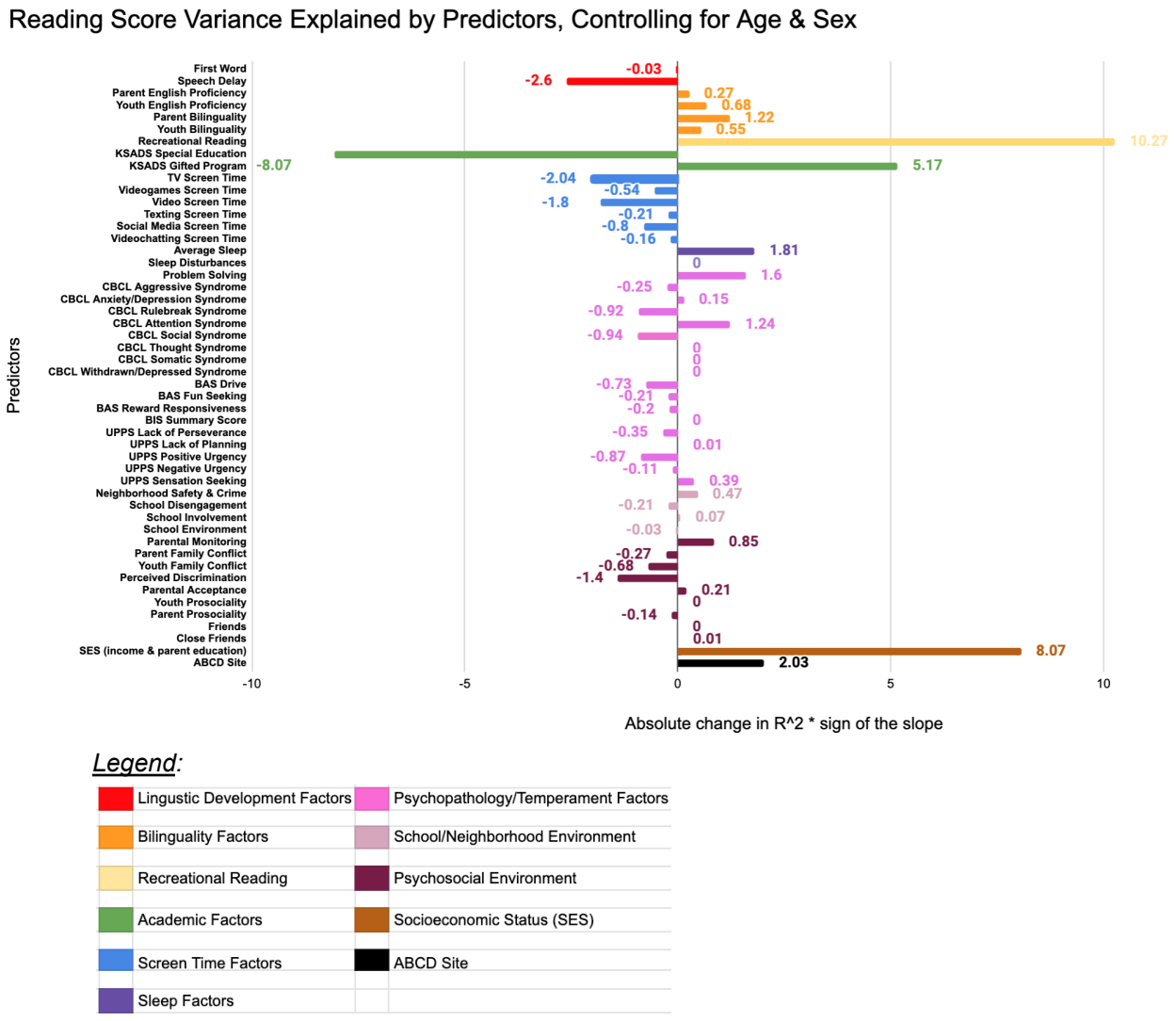


**Figure 2:** Reading Score Variance Explained by Predictors, Controlling for Age & Sex

The y-axis shows the predictors added to the linear regression model in which age and sex predict reading score. The x-axis represents the $\Delta R^2$ associated with each model multiplied by the sign of the slope between each variable and reading score. This shows the reading score variance explained by each predictor, above and beyond age and sex, and the direction of this relationship. The numeric labels to the right of each bar represent the $\Delta R^2$ when each of the predictors is added to the model in which age and sex predict reading score. Each bar is color-coded according to the behavioral domain it belongs to, indicated in the legend. [Refer to Methods for information on the different behavioral domains].

In the final stage of analysis, three sets of models were run to measure the $\Delta R^2$ when each domain of variables was added to the model with age and sex predicting reading score; the $\Delta R^2$ when each domain of variables was added to the model with age, sex, and SES predicting reading score; and the $\Delta R^2$ when each domain of variables was individually omitted from the full linear regression model in which all variables of interest and covariates predicted reading score. These $\Delta R^2$ values represent the variance in reading score explained controlling for age and sex (Figure 3; blue); the variance explained controlling for age, sex, and SES (Figure 3; green); and the unique variance explained controlling for all variables, respectively (Figure 3; yellow).

The Lingustic Development Factors account for 3.04% of variance in reading score when controlling for age and sex. The variance accounted for by the Linguistic Developmental Factors drops to 3% when controlling for age, sex, and SES and drops again to 0.7% when controlling for all variables. The Bilinguality Factors account for 3.42% of variance in reading score when controlling for age and sex. The variance accounted for by the Bilinguality Factors drops to 2.3% when controlling for age, sex, and SES and drops again to 1.1% when controlling for all variables. The Recreational Reading variable explains 9.06% of variance in reading score when controlling for age and sex. The variance explained by the Recreational Reading variable drops to 6.8% when controlling for age, sex, and SES and drops further to 3.4% when controlling for all variables. The Academic Factors explain 12.66% of variance in reading score when controlling for age and sex. The variance explained by the Academic Factors drops to 9.1% when controlling for age, sex, and SES and drops further to 4.1% when controlling for all variables.

The Screentime Factors account for 4.03% of variance in reading score when controlling for age and sex. The variance explained by the Screentime Variables drops to 1.3% when controlling for age, sex, and SES and then to 0.4% when controlling for all variables. The Sleep Factors account for 2.14% of the variance in reading score when controlling for age and sex. The variance explained by the Sleep Factors drops to 0.3% when controlling for age, sex, and SES and to 0.1% when controlling for all variables.

The Psychopathology/Temperament Factors account for 7.26% of the variance in reading score when controlling for age and sex. The variance explained by the Psychopathology/Temperament Factors drops to 3.7% when controlling for age, sex, and SES and drops further to 1.1% when controlling for all variables. The School/Neighborhood Environment Factors account for 1.1% of the variance in reading score when controlling for age and sex. The variance explained by the School/Neighborhood Environment Factors drops to 0.5% when controlling for age, sex, and SES and further to 0.1% when controlling for all variables. The Psychosocial Environment Factors explain 3.76% of the variance in reading score when controlling for age and sex. The variance explaind by the Psychosocial Environment Factors drops to 1.6% when controlling for age, sex, and SES and to 0.5% when controlling for all variables.

ABCD Site accounts for 1.99% of the variance in reading score when controlling for age and sex. The variance explained by ABCD Site drops to 1.7% when controlling for age, sex, and

SES and to 1.3% when controlling for all variables. SES accounts for 9.26% of the variance in reading score when controlling for age and sex, which drops to 2.2% when controlling for all variables. Lastly, age and sex explain 4.7% of the variance in reading score when controlling for all variables.

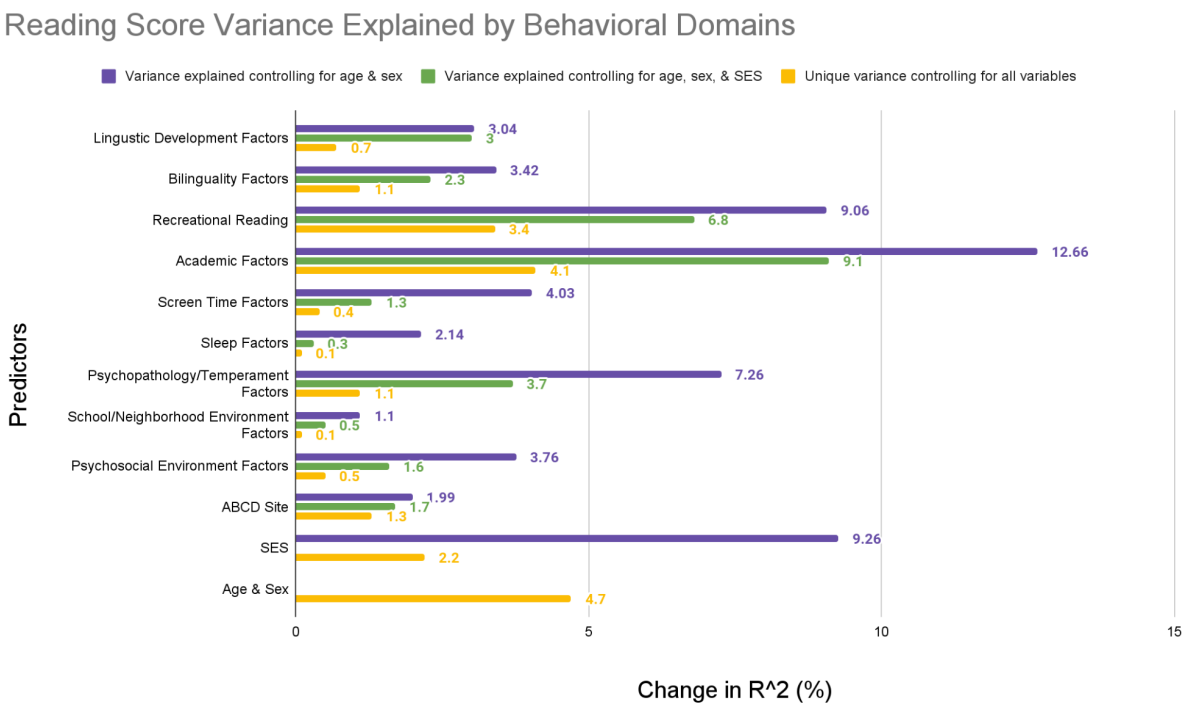


**Figure 3:** Reading Score Variance Explained by Behavioral Domains

The y-axis shows the sets of variables in behavioral domains included in three different linear regression models that predict reading score. The x-axis represents the $\Delta R^2$ associated with each model and variable domain. The numeric labels to the right of each bar represent the $\Delta R^2$ value associated with a given model and variable domain. The different colors of the bars represent the different models. The purple bars indicate the $\Delta R^2$ when each variable domain is added to the model with age and sex predicting reading score, which translates into the variance in reading score explained by each variable domain, controlling for age and sex. The green bars indicate the $\Delta R^2$ when each variable domain is added to the model with age, sex, and SES predicting reading score, which translates into the varaince in reading score explained by each domain group, controlling for age, sex, and SES. The yellow bars indicate the and the $\Delta R^2$ when each variable domain is individually omitted from the full linear regression model in which all variables of interest and covariates predict reading score, which translates into the unique variance in reading score explained by each variable domain, controlling for all other variables.

# Discussion

I ran three different sets of linear regression models to determine the unique and overlapping variability in reading score that different variables and behavioral domains accounted for. The first set of models were run to determine the amount of variability in reading score explained by three different models: one null model in which age and sex predicted reading score; another null model in which age, sex, and SES predicted reading score; and a full model in which all variables of interest predicted reading score. The second set of models were run to determine the variance in reading score explained by individual variables in these domains, controlling for age and sex. The third set of models were run to determine the variability in reading score explained by each behavioral domain and the extent to which these domains overlap with SES and each other in the variability they account for in reading score. Recreational Reading, Academic Factors, and SES yielded the largest effect sizes on their own and all of the behavioral domains were mediated by SES and each other, suggesting that variability in reading performance was shared amongst these domains.

*Variability in Reading Performance Explained by 3 Different Models*

By comparing the $R^2$ values associated with the three models from the first set of analyses, we can ascertain how much variance in reading score each model accounted for. The $R^2$ value increased from 0.0484 with age and sex as predictors to 0.141 with age, sex, and SES as predictors. This entails that SES on it's own accounted for a substantial amount of variance in reading score that was not captured by age and sex alone. The $R^2$ value increased again from 0.141 with age, sex, and SES as predictors to 0.336 with all variables of interest as predictors. This jump in $R^2$ value suggests that the remaining variables of interest accounted for a substantial amount of variance in reading score that was not captured by age, sex, and SES alone.

*Variability in Reading Scores Explained by Experiential Factors*

Experiential factors, defined as factors that explain how school-aged children spend their time, accounted for some variability in reading score. They may have also mediated some of the relationships observed between reading score and variables of interest.

Recreational reading yielded the largest effect size when controlling for age and sex, with a 0.1027 $\Delta R^2$, where higher reading scores were associated with more time spent reading for recreation. Children who are good at reading are motivated to read more, which entails growth in vocabulary and reading skills and widen the gap between good and bad readers over time (Clark, 2006). Another study discovered in their cohort of adolescents that reading for pleasure was more strongly associated with cognitive progress in math and vocabulary than parental education (Sullivan, 2015). Overall, the finding that recreational reading explained a significant amount of variance in reading score suggests that a child's intrinsic motivation to read may lead them to read more, which could then lead to them developing more reading skills.

The variability in reading explained by Recreational Reading overlapped with other behavioral domains in the model. The difference between the $\Delta R^2$ value predicted by Recreational Reading with and without controlling for all other variables of interest was 0.0566, which highlights the overlapping variability in reading score predicted by Recreational Reading and other behavioral domains. Based on the literature, I hypothesize that the variability in reading explained by Recreational Reading may overlap with Academic Factors and Screen Time Factors. Bobel found that children in gifted programs spent more time reading for pleasure than for information (1981). If children in gifted programs dedicate more time to reading, they may spend more time developing their reading skills, and therefore may have a higher reading performance. A potential mechanism to explain the overlapping variability between Recreational Reading and Screen Time Factors is that the more time children spend on screens, the less time they may spend reading for recreation. Wiecha found that children whose parents limit their television time spend more time reading (2001). Another study by Han found a negative association between the time children spent reading and the time they spent watching screens (2019). If children spend more time on screens and less time reading, they may dedicate less time to developing their reading skills and could have a poorer reading performance.

Both of the Academic Factors, KSADS Special Education and KSADS Gifted Program, yielded large effect sizes when controlling for age and sex. KSADS Special Education yielded a 0.0807 $\Delta R^2$, where lower reading scores were associated with enrollment in special education programs. KSADS Gifted Program yielded a 0.0517 $\Delta R^2$, where higher reading scores were associated with enrollment in gifted programs. I propose that the Gifted Program accounts for a substantial amount of variability in reading performance because children are initially admitted into such programs due to the fact that they perform at a higher level in school than their peers, including their reading performance. In one study, Rowe et al, reported that the test scores of adolescent participants in gifted and talented school programs on the Full Scale IQ (FSIQ), General Ability Index (GAI), and the Verbal Comprehension and Perceptual Reasoning exams were associated with their standardized test scores in reading comprehension (2012). Given that these children perform at a higher academic level, it follows that their reading scores would be higher than those of their peers. Once children are admitted into gifted programs, they are often given more resources and encouragement to succeed in reading. One case study found that students in gifted programs were initially frustrated when asked to read above their instructional level and struggled with doing so, but they were given extra encouragement from instructors and were eventually able to succeed in reading more advanced texts (Reis, 2009). Moreover, the finding that the Gifted Program explains a substantial amount of variability in reading performance suggests that perhaps children enter these programs with above average reading abilities and are then provided special resources and attention to foster these abilities through the program.

On the other hand, I hypothesize that the Special Education Programs account for a substantial amount of variability in reading performance because children are initially admitted into these programs due to performing at a lower academic level than their peers. As a result,

there are many children in these programs with learning disabilities that may impair their reading performance. Studies have found that almost 80% of children eligible for special education enrollment have serious reading problems, such as dyslexia (Dyslexia: What Brain Research Reveals About Reading). Additionally, in some cases there appears to be a lack of effective interventions to improve the reading performance of children enrolled in special education programs. One study found that most special education teachers seemed uncertain as to how to improve their students' reading comprehension skills and they often failed to use complicated strategies to achieve this (Klinger et al, 2010). The finding that special education programs explain a significant amount of variability in reading performance suggests that children in these programs may have pre-existing reading difficulties that make them eligible for enrollment in special education and they may not be given adequate instruction to improve their reading performance once enrolled in the programs.

The variability in reading explained by Academic Factors overlapped with other behavioral domains. The difference between the $\Delta R^2$ value predicted by Academic Factors with and without controlling for other behavioral domains was 0.0856. I propose that the variance explained by Academic Factors may overlap with Recreational Reading and Linguistic Development. The relationship between Academic Factors and Recreational Reading is highlighted above. A possible explanation for the overlapping variability between Academic Factors and Linguistic Development is that children who develop linguistic skills faster are potentially more likely to be enrolled in gifted programs. One case study found that a child enrolled in a gifted program developed language skills at a rapid pace (Hatice and Nuket, 2018). Another study found that early development of speech, movement, and reading is common in young children later enrolled in gifted programs and is related to high intellectual performance (Miraca, 1997). Such children may receive more advanced reading instruction and read at a more advanced level in the gifted programs, possibly leading them to have better reading performances.

The variability in reading explained by Screen Time Factors overlapped with other behavioral domains too. The difference between the $\Delta R^2$ predicted by Screen Time Factors with and without controlling for all other behavioral domains was 0.0363. I propose that the variance explained by Screen Time Factors overlaps with Sleep Factors. One possible mechanism to explain the ties between these factors is the negative relationship between adolescent screen time and sleep observed in the literature. One study found negative bidirectional relationships between sleep duration and media use in school-aged children (Magee et al, 2014). Additionally, screen time was negatively associated with sleep outcomes in 90% of the 67 studies reviewed by Hale and Guan (2015). These studies suggest that children who spend more time on screens get less sleep and poorer quality sleep. In another study, prolonged insufficient sleep negatively impacted cognitive and affective functioning (Short and Chee, 2019). Therefore, these children who engage in more screen time and get less sleep may be at risk of having poorer cognitive functioning, which could then put them at risk for poorer reading performance.

*Variability in Reading Scores Explained by SES and SES Mediation Effects*

SES accounted for some of the variability observed in reading score, yielding a large effect size of 0.0807 $\Delta R^2$ when controlling for sex, where higher reading scores were associated with higher SES. SES also appeared to serve as a mediator in the relationships between behavioral domains and reading score because there was an overlap in the variability in reading predicted by SES and each behavioral domain.

Sleep Factors had the largest SES mediation of all behavioral domains observed in this study. The difference between the $\Delta R^2$ predicted by Sleep Factors with and without controlling for SES was 0.0184, which highlights the overlapping variability in reading predicted by both SES and Sleep Factors. SES mediated almost all the variability in reading explained by Sleep Factors and this result is driven by the sleep disturbances measure. A potential explanation for this overlapping variability is that children from low SES backgrounds have a higher rate of sleep related issues, as documented in the literature. Low SES is associated with poor sleep quality, shorter sleep duration, and more daytime sleepiness in youths (Felden et al, 2015). Buckhalt found that many children from low SES backgrounds don't sleep well and that these sleep problems consequently have greater effects on their cognitive functioning and academic achievement (2011). Sleep is associated with intellectual capacity and academic performance in children (Buckhalt et al, 2009). Therefore, children from lower SES backgrounds may suffer greater tolls on their cognitive functioning as a result of their sleep related issues, and their impaired cognitive functioning may then lead to poorer reading performances.

The variability in reading explained by Recreational Reading also overlapped with SES. The difference between the $\Delta R^2$ value predicted by Recreational Reading with and without controlling for SES was 0.023, which highlights the overlapping variability in reading predicted by SES and Recreational Reading. A potential explanation for this overlapping variability is that children from lower SES backgrounds have fewer interactions with reading materials and activities relative to children from higher SES backgrounds. Sociocultural literature has posited that children from impoverished backgrounds have poorer literacy not as a direct result of their household income, but because they don't receive enough literacy experiences in their homes (Bhattacharya, 2010). Froiland additionally found that neighborhood SES predicted home literacy, defined as parents and children reading together and the number of books in a home, which subsequently predicted early literacy abilities in children (2013b). If children from low SES have fewer literacy opportunities in their homes, they may also have fewer opportunities to read for pleasure. Therefore, these children may have fewer opportunities to develop reading skills, and subsequently have poorer reading performance.

The variability in reading explained by Academic Factors overlapped with SES. The difference between the $\Delta R^2$ value predicted by Academic Factors with and without controlling for SES was 0.0356. A potential mechanism to explain the overlapping variability accounted for by SES and Academic Factors is the differential representation of children from high and low SES in gifted programs. McBee found that children from low SES were underrepresented in

gifted programs (2006). If there are fewer children from low SES backgrounds in gifted programs, they are potentially not given the same opportunities for advanced reading instruction as their peers from higher SES backgrounds. Schools in different districts may have differential access to gifted programs, so that schools with less money may not have the same resources for gifted programs as schools with more money. Therefore, the lack of representation of children from low SES backgrounds in these programs may be a direct result of the funding their school receives. Overall, the lack of representation of children from low SES backgrounds in gifted programs, whether or not it is a direct result of differences in school funding, may entail fewer opportunities for these children to advance their reading skills and potentially be tied to these children demonstrating poorer reading performance.

The variability in reading explained by Screen Time Factors overlaps with SES. The difference between the $\Delta R^2$ predicted by Screen Time Factors with and without controlling for SES is 0.0273. A potential explanation for the overlapping variability between Screen Time Factors and SES is that children from low SES backgrounds spend more time on screens. One study found that female children in low SES neighborhoods had significantly more screen time than those in high SES neighborhoods (Carson et al, 2010). Another study concluded that school-aged children from low SES backgrounds had more access to screens in their bedrooms (Tandon et al, 2012). One study also found a negative association between the time children spent reading and the time they spent watching screens (Han, 2019). Therefore, children who spend more time on screens may spend less time reading for pleasure. This relationship suggests that children from low SES backgrounds may spend more time engaged with screens and less time engaged with reading materials and that as a result, these children may have poorer reading performance due to a lack of reading practice.

The variability in reading explained by Psychopathology/Temperament Factors also overlapped with SES. The difference between the $\Delta R^2$ predicted by Psychopathology/Temperament Factors with and without controlling for SES was 0.0356. I hypothesize that the overlapping variability between Psychopathology/Temperament Factors and SES may be due to the overrepresentation of psychopathology in children from low SES backgrounds. In the literature, measures of different domains in psychopathology and temperament were positively or negatively associated with school-aged reading and literacy. Children from lower SES backgrounds have higher rates of psychopathology (Peverill et al, 2021). Additionally, one study found that adolescents who were impoverished since birth or who became impoverished between the ages of 0-11 years displayed increased conduct problems by age 15 (Anselmi, 2012). These children with psychopathology problems may then be more likely to have poor reading performance. Goldston et al. found that adolescents with poor reading skills had higher levels of ADHD, affective and anxiety disorders, and more functional impairment in many areas (2007). Another study found that harm avoidance and total anxiety problems were positively correlated with reading fluency in young female participants (Grills-Taquechel et al, 2012). Overall, if children from lower SES backgrounds have higher rates of psychopathology and if children with psychopathology problems demonstrate poorer reading performance, this

relationship may explain how SES and psychopathology are linked in their ability to explain variability in adolescent reading performance.

*Variability in Reading Score Explained by Developmental and Psychopathology Traits*

Speech Delay yielded one of the largest effect sizes when controlling for age and sex, with a 0.026 $\Delta R^2$, where higher reading scores were associated with less speech delay. I hypothesize that Speech Delay accounts for a significant amount of variability in reading score because the age at which a child begins to speak is related to the trajectory of their linguistic and reading development. One study that lends support to this theory showed that late talkers scored in the same range as their normally developing counterparts in reading skills at ages 6 and 7, but then had worse scores on the same tasks by ages 8 and 9, suggesting that slow early language development predisposes children for slower acquisition of language skills and lower performance in language skills into adolescence (Rescorla, 2002). Another study found that the age at which children acquire language skills has long term effects on their reading and language skills and on the neural pathways underlying these skills into their school-age years (Preston, 2010). Overall, the finding that speech delay explains a significant amount of variance in reading scores suggests children who speak later than average may have a slower linguistic developmental trajectory, and subsequently acquire reading skills at a slower pace, than children who speak earlier than average or on time.

The variability in reading explained by Psychopathology/Temperament Factors overlapped with other behavioral domains. The difference between the $\Delta R^2$ predicted by Psychopathology/Temperament Factors with and without controlling for other behavioral domains was 0.0616. I hypothesize that the variability explained by Psychopathology/Temperament Factors may overlap with Academic Factors and Screen Time Factors. Children in special education programs have more psychopathology problems that may predispose them to poorer reading performance. One study found that ODD, ADHD, parent-report internalizing behaviors, and teacher-reported externalizing behaviors were associated with special education enrollment for adolescents (Levine et al, 2013). Another countrywide survey found that half of the children in special education programs et the criteria for an ADHD diagnosis (Bussing et al, 2011). These findings show that psychopathology diagnoses are common for children enrolled in special education programs, particularly ADHD diagnoses. In the literature, ADHD is associated with reading issues. One study found that young adults with ADHD performed significantly worse than control groups on reading speed, response to questions, and a test of reading comprehension (Miranda et al, 2013). Another study confirmed that reading comprehension problems are common in adolescents with ADHD and found a significant relationship between sustained attention and reading comprehension in adolescents with and without ADHD (Pnina and Lilach, 2013). Overall, if children in special education programs are often diagnosed with ADHD and ADHD is associated with reading difficulties, these children may be likely to have poorer reading performance.

The variability in reading explained by Psychopathology/Temperament Factors may also overlap with Screen Time Factors. One study with a cohort of 3,826 adolescents reported a within-person association between social media and television consumption with depressive symptoms (Boers et al, 2019). Another study found that video game use was associated with anxiety and that video game and computer use were correlated with depression in children (Maras et al, 2015). As previously established, one study discovered a negative correlation between time spent reading and watching screens in children (Han, 2019). Therefore, if children with psychopathology problems spend more time engaged with screens, they may also spend less time engaged with reading materials. This lack of reading practice may then lead to poorer reading performance in this demographic.

*Conclusions and Future Directions*

I identified relationships that exist between oral reading scores and many other variables that describe the individual and environmental context of ABCD Study participants. The discovery of such relationships will ultimately contribute to a multidimensional understanding of the factors that underpin reading proficiency in 9-10 year old children. My analysis project has the potential to identify attributes in children, such as differences in linguistic development; English fluency levels; attitudes towards reading; academic programs; screen time habits; sleep behaviors; psychopathology; temperament; school and neighborhood environments; psychosocial environment; and socioeconomic status, that account for variability in their reading abilities.

Development is dynamic and as children develop over time, there are many external factors that can change or influence their developmental trajectories. Therefore, identifying factors associated with adolescent reading performance can help us understand the ways in which childhood literacy is related to different factors over time. Adolescent reading proficiency is highly predictive of later outcomes, such as economic stability, career success, and opportunities for higher education. Therefore, identifying factors associated with adolescent reading performance provides insight into potential interventions to improve adolescent reading performance and potentially improve their future outcomes.

# Acknowledgements

I would like to thank Dr. Terry Jernigan and Dr. Clare Palmer for giving me the opportunity to undertake this honors project and for their support throughout this entire process. I would not have been able to complete this project without your guidance, knowledge, and patience. I want to thank Dr. Douglas Nitz for his continued support and mentorship. I would also like to thank Dr. Marta Kutas for supervising my honors project and advising me as I created my honors thesis presentation. Being a part of this program and completing this thesis has been a

pleasure and a highlight of my undergraduate career at UCSD. I am very thankful for this opportunity.

# References


Alpana Bhattacharya (2010) Children and Adolescents From Poverty and Reading Development: A Research Review, Reading & Writing Quarterly, 26:2, 115-139, DOI: 10.1080/10573560903547445.

Anselmi L., Menezes A., Hallal P., Wehrmeister F., Goncalves H., Barros F. C., Murray J., Rohde L. A. (2012). Socioeconomic Changes and Adolescent Psychopathology in a Brazilian Birth Cohort Study. Journal of Adolescent Health. 51. S5-S10. https://doi.org/10.1016/j.jadohealth.2012.06.026.

Bobel, Joseph, Jr. (1981). The Attitudes and Preferences of Thirty-Eight Gifted Intermediate Grade Readers Toward Their Recreational Reading Habits and a Profile of Their Pre-School Reading Readiness. Institute of Education Sciences. ED200926.

Boers E, Afzali MH, Newton N, Conrod P. Association of Screen Time and Depression in Adolescence. JAMA Pediatr. 2019;173(9):853–859. doi:10.1001/jamapediatrics.2019.1759.

Buckhalt Joseph A. (2009). Concurrent and Longitudinal Relations Between Children's Sleep and Cognitive Functioning: The Moderating Role of Parent Education. 80. 875-892. https://doi.org/10.1111/j.1467-8624.2009.01303.x.

Buckhalt Joseph A. (2011). Insufficient Sleep and the Socioeconomic Status Achievement Gap. Child Development Perspectives. 5. 59-65. https://doi.org/10.1111/j.1750-8606.2010.00151.x.

Bussing R., Zima B. T., Perwien A. R., Belin T. R., Widawski M. (2011). Children in Special Education Programs: Attention Deficit Hyperactivity Disorder, Use of Services, and Unmet Needs. American Public Health Association.

Carson, V., Spence, J.C., Cutumisu, N. et al. Association between neighborhood socioeconomic status and screen time among pre-school children: a cross-sectional study. BMC Public Health 10, 367 (2010). https://doi.org/10.1186/1471-2458-10-367.

Clark, Christina & Rumbold, Kate. (2006). Reading for Pleasure: A Research Overview. Institute of Education Sciences. ED496343.

"Dyslexia: What Brain Research Reveals About Reading." Dyslexia: What Brain Research Reveals About Reading | LD Topics | LD OnLine, www.ldonline.org/article/10784/.

Esquierdo, Joy J & Arreguin-Anderson, Maria. (2012). The "Invisible" Gifted and Talented Bilingual Students: A Current Report on Enrollment in GT Programs. Journal for the Education of the Gifted. 35. 35-47. https://doi.org/10.1177/0162353211432041.

Froiland, John & Oros, Emily. (2013). Intrinsic motivation, perceived competence and classroom engagement as longitudinal predictors of adolescent reading achievement. Educational Psychology. 34. 119-132. 10.1080/01443410.2013.822964.

Froiland, John & Powell, Douglas & Diamond, Karen & Son, Seung-Hee. (2013). Neighborhood socioeconomic well-being, home literacy, and early literacy skills of at-risk preschoolers. Psychology in the Schools. 50. 755-769. 10.1002/pits.21711.

Goldston, B. D., Walsh, A., Mayfield Arnold, E., Palmes G., Snider E., & Wood F. (2007). Reading Problems, Psychiatric Disorders, and Functional Impairments from Mid- to Late Adolescence. Journal of the American Academy of Child & Adolescent Psychiatry. 46. 25-32. https://doi.org/10.1097/01.chi.0000242241.77302.f4.

Gomes Felden E P., Leite C. R., Rebelatto C. F., Andrade R. D., Beltrame T. S. (2015). Sleep in Adolescents of Different Socioeconomic Status: A Systematic Review. Revista Paulista de Pediatria. 33. https://doi.org/10.1016/j.rpped.2015.01.011.

Grills-Taquechel, A. E., Fletcher, J. M., Vaughn, S. R., & Stuebing, K. K. (2012). Anxiety and reading difficulties in early elementary school: evidence for unidirectional- or bi-directional relations?. Child psychiatry and human development, 43(1), 35–47. https://doi.org/10.1007/s10578-011-0246-1.

Hale, Lauren & Guan, Stanford. Screen Time and Sleep Among School-Aged Children and Adolescents: A Systematic Literature Review. Sleep Medicine Reviews. 21. 50-58. https://doi.org/10.1016/j.smrv.2014.07.007.

Han, Y. and Yan, H. (2019), Book or screen? a preliminary study on preschool children's reading and ICT using behaviors. Proceedings of the Association for Information Science and Technology, 56: 667-668. https://doi.org/10.1002/pra2.127.

Johanna L. Carpenter & Deborah A.G. Drabick (2011) Co-occurrence of linguistic and behavioural difficulties in early childhood: a developmental psychopathology perspective, Early Child Development and Care, 181:8, 1021-1045, DOI: 10.1080/03004430.2010.509795.

John Munro. (2005). The Learning Characteristics of Gifted Literacy Disabled Students. Gifted Education International. 19. 154-172. https://doi.org/10.1177/026142940501900209.

John Munro (2002) The Reading Characteristics of gifted literacy disabled students, Australian Journal of Learning Disabilities, 7:2, 4-12, DOI: 10.1080/19404150209546696.

Jonathan L. Preston, Stephen J. Frost, William Einar Mencl, Robert K. Fulbright, Nicole Landi, Elena Grigorenko, Leslie Jacobsen, Kenneth R. Pugh, Early and late talkers: school-age language, literacy and neurolinguistic differences, Brain, Volume 133, Issue 8, August 2010, Pages 2185–2195, https://doi.org/10.1093/brain/awq163.

Kadioglu Ates, Hatice & Afat, Nuket. (2018). A Case Study Investigating the Language Development Process, Early Literacy Experiences and Educational Problems of a Gifted Child. Journal for the Education of Gifted Young Scientists. 6. 36071.i.

Kathryn Lindholm-Leary (2014) Bilingual and biliteracy skills in young Spanish-speaking low-SES children: impact of instructional language and primary language proficiency, International Journal of Bilingual Education and Bilingualism, 17:2, 144-159, DOI: 10.1080/13670050.2013.866625.

Klinger J. K., Urbach J., Golos D., Brownwell M., Menon S. (2010). Teaching Reading in the 21st Century: A Glimpse at How Special Education Teachers Promote Reading Comprehension. Learning Disability Quarterly. https://doi.org/10.1177/073194871003300201.

Levine T. P., Lester B., Lagasse L., Shankaran S., Bada H. S., Bauer C. R., Whitaker T. M., Higgins R., Hammond J., Roberts M. B. (2013). Psychopathology and Special Education Enrollment in Children with Prenatal Cocaine Exposure. Journal of Developmental Behavioral Pediatrics. 33. 377-386. 10.1097/DBP.0b013e3182560cd9.

Magee CA, Lee JK, Vella SA. Bidirectional Relationships Between Sleep Duration and Screen Time in Early Childhood. JAMA Pediatr. 2014;168(5):465–470. doi:10.1001/jamapediatrics.2013.4183.

Maras D., Flament M. F., Murray M., Buchholz A., Henderson K. A., Obeid N., Goldfield G. S. (2015). Screen time is Associated with Depression and Anxiety in Canadian Youth. Preventive Medicine. 73. 133-138. https://doi.org/10.1016/j.ypmed.2015.01.029.

Matthew Peverill, Melanie A. Dirks, Tomás Narvaja, Kate L. Herts, Jonathan S. Comer, Katie A. McLaughlin. (2021). Socioeconomic status and child psychopathology in the United States: A meta-analysis of population-based studies. Clinical Psychology Review. 83. https://doi.org/10.1016/j.cpr.2020.101933.

McBee, Matthew T. (2006). A Descriptive Analysis of Referral Sources for Gifted Identification Screening by Race and Socioeconomic Status. Journal of Advanced Academics. 17. 103-111. https://doi.org/10.4219/jsge-2006-686.

Miraca U.M. Gross (1999) Small poppies: Highly gifted children in the early years, Roeper Review, 21:3, 207-214, DOI: 10.1080/02783199909553963.

Miranda A., Mercader J., Inmaculada F., Colomer C. (2013). Reading Performance of Young Adults with ADHD Diagnosed in Childhood: Relations with Executive Functioning. Journal of Attention Disorders. https://doi.org/10.1177/1087054713507977.

Ornoy, A., Uriel, L., & Tennenbaum, A. (1993). Inattention, hyperactivity and speech delay at 2–4 years of age as a predictor for ADD-ADHD syndrome. Israel Journal of Psychiatry and Related Sciences, 30(3), 155–163.

Pnina, Stern & Shalev, Lilach. (2013). The Role of Sustained Attention and Display Medium in Reading Comprehension Among Adolescents with ADHD and Without It. Research in Developmental Disabilities. 34. 431-439. https://doi.org/10.1016/j.ridd.2012.08.021.

Reis, Sally M & Boeve, Hope. (2009). How Academically Gifted Elementary, Urban Students Respond to Challenge in an Enriched, Differentiated Reading Program. Journal for the Education of the Gifted. https://doi.org/10.1177/016235320903300204.

Rescorla. Leslie. (2005).Age 13 Language and Reading Outcomes in Late-Talking Toddlers. Journal of Speech, Language, and Hearing Research. 48. 459-472. https://doi.org/10.1044/1092-4388(2005/031).

Rescorla. Leslie. (2002). Language and Reading Outcomes to Age 9 in Late-Talking Toddlers. Journal of Speech, Language, and Hearing Research. https://doi.org/10.1044/1092-4388(2002/028)

Rowe, E. W., Miller, C., Ebenstein, L. A., & Thompson, D. F. (2012). Cognitive predictors of reading and math achievement among gifted referrals. School Psychology Quarterly, 27(3), 144–153. https://doi.org/10.1037/a0029941.

Sara Anderson, Tama Leventhal & Véronique Dupéré (2014) Exposure to Neighborhood Affluence and Poverty in Childhood and Adolescence and Academic Achievement and Behavior, Applied Developmental Science, 18:3, 123-138, DOI: 10.1080/10888691.2014.924355.

Short, Michael A & Chee, Michael W. L. (2019). Chapter 3 - Adolescent Sleep Restriction Effects on Cognition and Mood. Progress in Brain Research. 246. 55-71. https://doi.org/10.1016/bs.pbr.2019.02.008.

Sullivan, A. and Brown, M. (2015), Reading for pleasure and progress in vocabulary and mathematics. Br Educ Res J, 41: 971-991. https://doi.org/10.1002/berj.3180.

Tandon, P.S., Zhou, C., Sallis, J.F. et al. Home environment relationships with children's physical activity, sedentary time, and screen time by socioeconomic status. Int J Behav Nutr Phys Act 9, 88 (2012). https://doi.org/10.1186/1479-5868-9-88.

Uccelli, Paola & Paez Mariela. (2007). Narrative and Vocabulary Development of Bilingual Children From Kindergarten to First Grade: Developmental Changes and Associations Among English and Spanish Skills. Language, Speech, and Hearing Services in Schools. 38. 225-236. https://doi.org/10.1044/0161-1461(2007/024).

van Zwieten, A., Teixeira-Pinto, A., Lah, S., Nassar, N., Craig, J. C., & Wong, G. (2020). Socioeconomic Status During Childhood and Academic Achievement in Secondary School. Academic pediatrics, S1876-2859(20)30570-2. Advance online publication. https://doi.org/10.1016/j.acap.2020.10.013.

Wiecha J. L., Sobol A. M., Peterson K. E., Gortmaker S. L. (2001). Household Television Access: Associations With Screen Time, Reading, and Homework Among Youth. Ambulatory Pediatrics. 1. 244-251. https://doi.org/10.1367/1539-4409(2001)001<0244:HTAAWS>2.0.CO;2.

Willms, Douglas J. (2003). Literacy Proficiency of Youth: Evidence of Converging Socioeconomic Gradients. International Journal of Educational Research. 39. 247-252. https://doi.org/10.1016/j.ijer.2004.04.005.

Zebroff, D., Kaufman, D. Texting, reading, and other daily habits associated with adolescents' literacy levels. Educ Inf Technol 22, 2197–2216 (2017). https://doi.org/10.1007/s10639-016-9544-3.

# Appendix

## Appendix A: Variable Table

| Variable Name | Domain | Definition | Item or Summary | Changes Applied | Valid Range of Values |
|---|---|---|---|---|---|
| Reading Score | | NIH Toolbox Oral Reading Recognition Test Score | Item | | 59-180 |
| First Word | Linguistic Development Factors | (Parent report) Age (in months) child was able to say their first word | Item | | 0-60 |
| Speech Delay | Linguistic Development Factors | (Parent report) Whether the child's speech development was earlier, average, or later than most children | Item | Numeric transformation | 1-5 |
| Parent English Proficiency | Bilinguality Factors | (Parent report) How well the participant's parent speaks English | Item | Numeric transformation | 1-4 |
| Youth English Proficiency | Bilinguality Factors | (Youth report) How well the child speaks English | Item | Numeric transformation | 1-4 |
| Parent Bilinguality | Bilinguality Factors | (Parent report) Parent speaks another language other than English | Item | Numeric transformation | 0-1 |
| Youth Bilinguality | Bilinguality Factors | (Youth report) Child speaks another language other than English | Item | Numeric transformation | 0-1 |
| Recreational Reading | Recreational Reading | (Parent report) Hours per week child spends reading for pleasure | Item | Removed values >30 because any value above 30 is highly unlikely. 77 values outside valid range were omitted. | 0-30 |
| KSADS Special Education | Academic Factors | (Parent report) Whether or not a child | Summary | Recoded so that any value >= 1 is a 1 | 0-1 |

| | | | | | |
|---|---|---|---|---|---|
| | | receives special education services | | | |
| KSADS Gifted Program | Academic Factors | (Parent report) Whether the parent's child is a member of their school's Gifted Program | Item | Numeric transformation | 0-1 |
| TV Screen Time | Screen Time Factors | (Youth report) Hours spent watching TV per week | Item | Numeric transformation | 1-7 |
| Videos Screen Time | Screen Time Factors | (Youth report) Hours spent watching videos (eg YouTube) per week | Item | Numeric transformation | 1-7 |
| Videogames Screen Time | Screen Time Factors | (Youth report) Hours spent playing videogames per week | Item | Numeric transformation | 1-7 |
| Texting Screen Time | Screen Time Factors | (Youth report) Hours spent texting per week | Item | Numeric transformation | 1-7 |
| Social Media Screen Time | Screen Time Factors | (Youth report) Hours spent on social media per week | Item | Numeric transformation | 1-7 |
| Videochatting Screen Time | Screen Time Factors | (Youth report) Hours spent videochatting per week | Item | Numeric transformation | 1-7 |
| Average Sleep | Sleep Factors | (Parent report) Hours of sleep child gets most nights | Item | | |
| Sleep Disturbances | Sleep Factors | (Parent report) Sum of sleep disorders & disturbances a child presents | Summary | | 26-126 |
| CBCL Aggressive Syndrome | Psychopathology Factors | (Parent report) Sum of Aggressive CBCL Syndrome Scale | Summary | | 0-36 |
| CBCL Anxiety/Depression Syndrome | Psychopathology Factors | (Parent report) Sum of Anxiety/Depression Syndrome Scale | Summary | | 0-26 |

| | | | | | |
|---|---|---|---|---|---|
| CBCL Rulebreak Syndrome | Psychopathology Factors | (Parent report) Sum of Rulebreak Syndrome Scale | Summary | | 0-23 |
| CBCL Attention Syndrome | Psychopathology Factors | (Parent report) Sum of Attention Syndrome Scale | Summary | | 0-20 |
| CBCL Social Syndrome | Psychopathology Factors | (Parent report) Sum of Social Syndrome Scale | Summary | | 0-19 |
| CBCL Thought Syndrome | Psychopathology Factors | (Parent report) Sum of Thought Syndrome Scale | Summary | | 0-22 |
| CBCL Somatic Syndrome | Psychopathology Factors | (Parent report) Sum of Somatic Syndrome Scale | Summary | | 0-18 |
| CBCL Withdrawn/Depressed Syndrome | Psychopathology Factors | (Parent report) Sum of Withdrawn/Depressed Syndrome Scale | Summary | | 0-16 |
| Problem Solving | Problem Solving | (Youth report) Will's Problem Solving summary score | Summary | Numeric transformation | 6-30 |
| BAS Drive | Temperament Factors | (Youth report) BAS Drive summary score | Summary | | 0-12 |
| BAS Fun Seeking | Temperament Factors | (Youth report) BAS Fun Seeking summary score | Summary | | 0-12 |
| BAS Reward Responsiveness | Temperament Factors | (Youth report) BAS Reward Responsiveness summary score | Summary | | 0-15 |
| BIS Summary Score | Temperament Factors | (Youth report) BIS summary score | Summary | | 0-21 |
| UPPS Lack of Perseverance | Temperament Factors | (Youth report) Lack of perseverance summary score | Summary | | 4-16 |

| | | | | | |
|---|---|---|---|---|---|
| UPPS Lack of Planning | Temperament Factors | (Youth report) Lack of planning summary score | Summary | | 4-16 |
| UPPS Positive Urgency | Temperament Factors | (Youth report) Positive urgency summary score | Summary | | 4-16 |
| UPPS Negative Urgency | Temperament Factors | (Youth report) Negative urgency summary score | Summary | | 4-16 |
| UPPS Sensation Seeking | Temperament Factors | (Youth report) Sensation seeking summary score | Summary | | 4-16 |
| Neighborhood Safety & Crime | Neighborhood Environment | (Parent report) How safe/unsafe parent feels in their neighborhood | Summary | Numeric transformation | 1-5 |
| School Disengagement | School Environment | (Youth report) Child's level of disengagement in school | Summary | Numeric transformation | 2-8 |
| School Involvement | School Environment | (Youth report) Child's level of involvement in school | Summary | Numeric transformation | 4-16 |
| School Environment | School Environment | (Youth report) Child's perceived quality of school environment | Summary | Numeric transformation | 6-24 |
| Parental Monitoring | Psychosocial Environment | (Youth report) The average value of all parental monitoring (pm) variables | Summary | Numeric transformation | 1-5 |
| Parent Family Conflict | Psychosocial Environment | (Parent report) Level of family conflict | Summary | | 0-9 |
| Youth Family Conflict | Psychosocial Environment | (Youth report) Level of family conflict | Summary | | 0-9 |
| Perceived Discrimination | Psychosocial Environment | (Youth report) degree of perceived discrimination on the | Summary | Numeric transformation | 1-5 |

| | | | | | |
|---|---|---|---|---|---|
| | | basis of youth's race & ethnicity | | | |
| Parental Acceptance | Psychosocial Environment | (Youth report) degree of perceived acceptance & care from both parents | Summary | | 1-3 |
| Youth prosociality | Psychosocial Environment | (Youth report) prosocial behavior self report | Summary | Numeric transformation | 0-2 |
| Parent Prosociality | Psychosocial Environment | (Parent report) prosocial behavior self report | Summary | Numeric transformation | 0-2 |
| Friends | Psychosocial Environment | (Youth report) Total number of child's friends | Item | Quantile normalization transformation | 0-200 |
| Close Friends | Psychosocial Environment | (Youth report) Total number of child's close friends | Item | Quantile normalization transformation | 0-200 |
| Combined Family Income | SES | (Parent report) Total combined family income over the past year | Item | Log transformation | 2500 - 250000 |
| Highest Level of Parent Education | SES | Highest level of education parents' education | Item | | < HS Diploma; HS Diploma/GED; some college; Bachelor; Post Graduate Degree |
| ABCD Site | ABCD Site | Site at which participant's data was collected | Item | Gave each site a valid name, such as changing "site01" to "CHLA" | |
| Age | Age | Age of child (in months) | Item | | 107-165 |
| Sex | Sex | Sex of child at birth | Item | | Male; Female |