

Teacher-Child Interaction and Child Outcomes:
Puget Sound Educational Service District Head Start Preschool Programs 2012-2013

Sarah Kelley

A Capstone project presented in partial fulfillment of the
requirements for the degree of
Master of Arts in Policy Studies
Interdisciplinary Arts and Sciences

Abstract

There is an achievement gap between high and low income preschool aged children. Head Start and state funded targeted preschool programs seek to decrease the achievement gap. Higher quality teacher-child interactions are associated with positive child outcomes in empirical research. The purpose of this study is to identify the relationship between teacher-child interaction scores and children's developmental gains over one academic year in Puget Sound Educational Service District's (PSESD) targeted preschool programs in Washington State. This study utilizes data from fall and spring 2012-13 with a sample of 120 classrooms and 1698 students from PSESD. Classroom Assessment Scoring System CLASS measures teacher-child interaction and a comprehensive observational checkpoint assessment measures child development. The major findings include low positive associations between classroom organization and physical and language development, and low positive associations between classroom instruction and cognitive and social emotional development. Emotional support has a low negative association to all areas of development. The modest positive relationships, consistent with previous literature, indicate CLASS can be used as a professional development and assessment tool to impact student outcomes. The negative relationships between emotional support and child outcomes are inconsistent with the literature. However in this study the negative relationship is present across all developmental domains indicating there could be validity issues with CLASS in emotional support scores or the process in which CLASS is administered. Further research to explore best practices in the implementation and use of CLASS is recommended.

Chapter 1- Purpose of the Study

Problem Statement

There is an achievement gap between higher and lower income children in school readiness. Young children ages 3-6 living in poverty are less likely to be able to write their name, recognize letters, and count to 20 plus than children living in non-poor households (Department of Education, 2010). Kindergarten readiness, academic and non-academic skills are correlated to math and reading achievement through fifth grade (Vi-Nhuan, L., 2006). Federal and state funded targeted preschool programs for low income families aim to reduce the achievement gap in school readiness and future academic achievement. Income eligible Head Start and state funded programs modeled after or combined with Head Start (e.g. Early Childhood Education and Assistance ECEAP in Washington State) provide education, family advocacy, nutritional and health services to 3-6 year old children. One strategy to increase Head Start's effectiveness at addressing this authorizing program mission of reducing the kindergarten readiness gap is to improve teaching practices in Head Start programs. The Office of Head Start has adopted Classroom Assessment Scoring System (CLASS) to monitor teacher-child interaction quality and provide resources that align with the assessment tool to promote evidence based effective teaching practices. This study identifies correlations between quality of teaching practices and child outcomes.

Head Start participation has been shown to impact children's social-emotional and cognitive development when compared to children in low-income households who did not receive Head Start services (Puma, et al, 2010). Higher quality classrooms have been empirically found to have greater impact on child outcomes. The quality of teacher-child interaction has greater impact on child outcomes than structure quality measures (Sabol, et al., 2013). Head Start has federally mandated an assessment of teacher-child interaction quality and mandated increases in teacher educational attainment. Teacher education is inconsistent in associations to child outcomes and in some studies higher teacher educational attainment is not associated to higher classroom quality (Early, 2007).

The majority of modern literature in early childhood uses CLASS to measure teacher-child interaction quality. Child assessments vary and studies usually use separate methods to measure different child development domains. This study will use a comprehensive measure of child development that is the same assessment Washington State is implementing in kindergarten classrooms (SB5427, 2011-12). The significance of this research is to identify relationships between teacher-child interaction quality and child outcomes on a comprehensive assessment. The reader will benefit from this approach because child outcome results will be comparable across developmental domains. Washington State will benefit from this study as a way to compare two assessment components of current quality improvement efforts and kindergarten readiness efforts. National audiences will benefit from this analysis to compare their own early childhood systems to Washington's current programs.

Research Purpose

The purpose of this explanatory study is to investigate how teacher-child interaction quality relates to school readiness in low income targeted Head Start preschool programs in Puget Sound Educational Service District (PSESD). The significance of this inquiry is to test the theory of teaching through interaction developed by Drs. Robert Pianta and Bridget Hamre, who developed the CLASS assessment. The independent variable teacher-child interaction quality will be defined as Classroom Assessment Scoring System CLASS scores measuring emotional support, classroom organization and instructional support. The dependent variable child outcomes will be defined as scores on Teaching Strategies GOLD, a teacher rated developmental checklist measuring children's development on 38 objectives in 6 areas, social emotional, language, literacy, cognitive, math and physical development. The control variable will be defined as participation in PSESD Head Start income eligible targeted programs. This study is a cross-sectional design that measures the differences between classrooms on quality and outcomes in one academic year. The research question is does quality of teacher-child interactions relate to child developmental outcomes in targeted preschool programs?

Chapter 2- Literature Review

The importance of early learning has gained increasing attention as research demonstrates how early childhood education can impact child outcomes. Kindergarten readiness, closing the achievement gap and supporting at risk young children are the cornerstones of targeted pre-kindergarten programs. Stakeholders and policy makers are interested in highest yielding investments in early childhood education and care. This literature review focuses on two Federal policies aimed at increasing quality in Head Start programs, teacher education requirements and teacher-child interaction assessment. Head Start program impacts and classroom quality are presented. Next evidence of the impact of teacher education and teacher-child interaction on child outcomes is reviewed. Finally improving the quality of teacher-child interactions is discussed

Head Start Program Impacts

The purpose of Head Start, established in 1965, is to increase school readiness in low-income three and four year olds. Key program domains are children's cognitive and social emotional development. Head Start participation has been empirically linked with positive social emotional and cognitive outcomes among low income three and four years when compared to a control group of low-income children not participating in Head Start programs. Due to feasibility and ethical reasons children in the control group are not denied access to preschool programs. In a nationally representative sample 60% of the control group children enrolled in some kind of early education program at some time in the first study year (13.8% of three year olds and 17.8% of the four years olds enrolled in Head Start programs). Despite more than half of the control group receiving some form of education and care, the study observed statistically significant Head Start impacts for three and four year olds enrolled in fall 2002. The four year old cohort outcomes were significant in language and literacy skills. The three year old cohort outcomes were significant in social emotional, literacy, language and math skills. In the following Head Start year the three year old cohort achieved significant gains in social skills and positive approaches to learning (Puma, et al, 2010).

Head Start Accountability and Quality

These findings lead to important policy questions; what is it about Head Start programs or private preschool programs that impacts positive child outcomes and how is that different than other education and care programs? How can we hold programs accountable for these aspects that influence child outcomes? The current political attention and federal and state dollars contributing to targeted preschool programs have generated interests in accountability. However the applicability of child outcome measures to program accountability is limited, the reliability and validity of standardized tests are lower with young children. This is attributed to their knowledge and capabilities being both unstable and contextual (La Paro, et al., 2004). There is a significant body of research that seeks to answer these questions by measuring classroom quality in relation to child outcomes.

Classroom quality is often measured with observational tools, there are more than 50 observational assessment tools applicable to early childhood settings. The challenges in using these tools are reliability between raters, validity of the tools in a variety of settings, frequency of the data collection and intended use of the data. Observational data has been used in both low and high stakes decision making. Low being program decisions such as professional development to pursue. High stakes being resource allocation, closing programs or requiring competition for grant in which the program might lose their funding (Ackerman, 2014).

Quality is typically measured by process or structural program features. Structural features include, adequacy of materials, space, safety, adult-child ratio, staff qualifications, wages and curriculum. Process measures are children's experience, classroom climate, teacher-child interactions and relationships. In the past decade a trend towards assessment and accountability using mixed measures of process and structural features has emerged (Downer, 2012). Meta-analysis of quality measures in early childhood settings found process measures to be more predictive of child outcomes than structural measures (Sabol, et al., 2013).

Federal Head Start policy makers have mandated “By September 30, 2013 at least 50% of Head Start teachers nation-wide must have a baccalaureate or advanced degree in Early Childhood Education... (Early Childhood Learning and Knowledge Center ECLKC, Head Start Statutory Degree).” Head Start implemented the use of Classroom Assessment Scoring System (CLASS), an observational teacher-child interaction quality assessment, at the grantee level. Effective in December 2011 programs falling below a minimum threshold or within the lowest 10% of nationwide CLASS scores are required to compete for their grant (ECLKC, Report to Congress).

Teacher Education

Existing research offers mixed results regarding the impact of staff education on measured classroom quality and child outcomes. In a national sample of privately funded childcare centers teacher education is reported to be associated to warmth, sensitivity, higher responsive interaction and positive child engagement in the classroom (Howes, 1997). Conversely several secondary analysis studies of the National Center for Early Development and Learning’s (NCEDL) Multi-State Study of Pre-kindergarten identified few or weak associations between teacher education, classroom quality and child outcomes. Using six teacher credential variables (years of education, degree, BA vs. no BA, major, state teaching certificate and CDA) there was no significant association to classroom quality measures. The only significant association to child outcomes was teacher’s years of education and children’s math skills (Early, et al., 2006). A modest association between BA-holding teachers and classroom quality was reported by Pianta et al. in a 2005 study. Researchers noted the presence of specialized training in early childhood had a stronger impact on classroom quality measures than the BA-holding teachers alone (Pianta, et al., 2005). In another secondary data analysis using seven studies, two studies reported BA or higher degree was associated to higher classroom quality, one study reported BA or higher degree had a negative association to classroom quality and four studies reported no association (Early, 2007). These varied results indicate the complex nature of relationships between teacher education and observed classroom quality. College education programs vary across universities within the same field. Early childhood education BA degree

programs teach a variety of different concepts, curriculum and competencies. Weak and inconsistent correlation between teacher education and classroom quality is problematic in tying quality improvement efforts in early childhood setting to increasing educational attainment of teachers. Instead of relying on teacher education alone to increase classroom quality more quality rating and improvement systems (QRIS) are being developed and implemented.

Many states are turning to market based QRIS to measure quality, publicize quality ratings and inform on quality improvement efforts. Aggregating quality measures across several states' QRIS researchers' applied the aggregated rating scale to the NCEDL multistate study of prekindergarten and the State-Wide Early Education Programs (SWEEP) Study. Researchers found the strongest quality predictor of child outcomes is teacher-child interaction CLASS scores and learning environment. Teacher qualifications and other structural program features were again found to be only a modest predictor of child outcomes. When all QRIS quality rating measures are aggregated the ratings do not consistently relate to child outcomes (Sabol, et al., 2013). The policy implications are to focus quality improvements on process and teacher-child interaction improvements. Although teacher education is not consistently predictive of classroom quality or child outcomes, early childhood higher education programs, state certificate programs and professional development training for teachers can focus on increasing interaction skills.

Teacher-child interaction

There are many studies that link quality teacher-child interactions in pre-kindergarten to positive child social-emotional and cognitive outcomes with varying levels of association strengths (Burchinal, et al., 2011, Moiduddin, et al., 2012, Gosse, et al., 2012 Zaslou, et al., 2010, Howes, et al., 2008, Mashburn, et al., 2008, Early, et al., 2005, Sabol, et al., 2013, Brophy-herb, et al., 2007, NICHD, 2005, NHCID, 2006, & Helburn, 1995). In the following section major finding will be reported for social and cognitive outcomes related to teacher-child interaction.

Classroom assessment scoring system CLASS is divided into three domains, instructional support, classroom organization and emotional support. Studies in the U.S. have consistently found classrooms score higher on average in classroom organization and emotional support and lower on instructional support (Burchinal, et al., 2011). A pattern of logical connections between social-emotional and cognitive child outcomes with the three domains of quality in CLASS is common across many studies. Classroom climate is related to both cognitive and social-emotional child outcomes. Emotional support is related to social-emotional child outcomes. Instructional support is related to cognitive child outcomes. (Moiduddin, et al., 2012). FACES 2009, conducted with a nationally representative Head Start sample found significant positive linear associations between instructional support which includes teacher language modeling and children's language and literacy skills. Classroom organization had a positive linear association to children's social skills. Non-linear associations, such that the associations were higher in higher quality classrooms, were identified between teacher instructional support and children's language skills. There was non-linear association between classroom positive climate and children's problem behavior (Moiduddin, E., et al., 2012).

Previous studies report non-linear relationships between classroom quality and child outcome measures indicating the impact of quality is only significant in a certain range. Under or over quality thresholds the slope is flat or curved (Zaslow, M., et al., 2010). Further research might better our understandings of the range in which quality impacts outcomes to inform effective quality improvement efforts.

NCEDL & SWEEP data was used to estimate quality of teacher-child interactions (CLASS measure) on cognitive and social outcomes. Instructional interactions were significantly correlated to vocabulary, rhyming, applied problems and letter naming. Emotional interactions were positively associated with children's social competence and negatively associated with problem behavior (Mashburn, A. J., 2008). Also using NCEDL & SWEEP Howes & Burchinal (2008) created a process quality composite measure. The process quality composite was correlated with literacy, language and teacher-child relationships were correlated with social and math skills. Burchinal identified threshold

effects in quality of teacher-child interaction. Teacher-child interaction had a higher association to social competence and less problem behavior in overall higher quality classrooms (2010). Using original data teacher reported negative child behavior was predictive of negative classroom climate and conversely reports of positive child behavior were predictive of positive classroom climate (Brophy-herb, 2007).

NICHD early child care effect size study in 2006 found higher quality care has moderate significant correlation on all cognitive and most social measures, as is consistent with finding from NICHD ECCYC 1998, 2000, 2001 and 2002. Many studies document the short term child outcomes and there is some evidence child gains predicted by teacher-child interaction quality are maintained through third grade, although the effects are diminishing over time. Teacher reported closeness to children, operationalized as teacher-child relationships was predictive of high standardized language and math test scores in 1st and 3rd grade and high teacher ratings on cognitive and attention skills measured in kindergarten and 2nd grade (NICHD, 2005, Peisner-Feinberg, et al., 2001, Helburn, S. W., Ed., 1995).

It is beneficial to the early childhood field to have the large scale longitudinal data sets such as NICHD and NCEDL & SWEEP data. As reviewed, there is significant evidence linking classroom quality to child outcomes and stronger evidence linking process quality measures to child outcomes. Most studies employ a variety of child assessments to measure child outcomes including standardized texts, interviews, and teacher and parent reports. The methodological rigor of studies measuring child outcomes would benefit from a comprehensive child assessment including all developmental domains in which raters are trained for inter-rater reliability. A comprehensive assessment would increase the construct validity in child outcome research and increase the comparability of results.

Improving teacher-child interaction quality

In a randomized control group experiment with 440 early childhood teachers, half were assigned to a 14 week course on CLASS teacher-child interaction. Teachers reported greater understanding of quality interaction and feelings of intentional teaching. In

observational assessment of teachers, program participants in the course training demonstrated higher quality emotional support and instructional support. The course was equally effective among teachers with BA and less than AA degrees (Hamre, et al. 2012). Another randomized control study assigned Head Start teachers to intensive professional development or regularly provided Head Start professional development. The experimental group at the end of the academic year had higher CLASS scores and environmental ratings scores. Children in the experimental group classrooms scored higher on language and pre-literacy measures. These experiments provide evidence supporting policy aimed at increasing teacher's interaction skills regardless of educational attainment. As mentioned earlier, educational attainment in early childhood related degrees that include interaction curriculum might also impact quality of interaction.

Conclusion

Head Start and early childhood programs are seeking to increase quality and impact child outcomes. Teacher education has small and inconsistent associations to positive child outcomes. Teacher-child interactions are moderately and consistently associated to positive outcomes and more predictive of outcomes than structural quality measures. Establishing thresholds for quality impacts on child outcomes could be a useful in focusing quality improvement resources. Utilizing a comprehensive child assessment measure for research could increase comparability of study results. Professional development aimed at training teachers in relational and interaction skills is a promising method to impact child outcomes and school readiness.

Chapter 3- Methodology

Research Question & Hypothesis

Does quality of teacher-child interactions relate to child developmental gains in targeted preschool programs? Children in classrooms with higher quality teacher-child interactions are predicted to have greater developmental gains from fall to spring in the 2012-2013 academic years. This explanatory research project seeks to address an aspect of what influences developmental gains in targeted preschool programs with the expectation that teacher quality relates to child gains. This is a theory testing model in which theory of teacher-child relationships (Downer, J., 2010) is tested with observations of teacher quality and child outcomes.

Data Collection

The population studied is low income targeted Head Start students under the Puget Sound Educational Service District PSESD located in Washington State. The data is from the 2012-2013 school year in which PSESD collected data from 120 classrooms serving 1698 Head Start (low income) students. The variables are operationalized with two observational assessment tools measuring teacher quality and student outcomes.

Population

This sample included 120 classrooms and 1698 students, missing data is excluded list wise. There was an even distribution for gender, 849 females and 849 males. A little over half the students speak English (55%), a third speak Spanish (30%) and a small percentage speak African (5%). All other languages represent less than 5% of the student population (see appendix table and graph E.1). The student population consisted of 43% Caucasian, 18% African American, 16% Multiracial and 12% American Indian or Alaskan Native (see table and pie chart appendix E.2). In the spring nearly three quarters (72%) of the students are prekindergarten 4 year olds and a quarter (27%) are preschool 3 year olds (see appendix table E.3).

Research Design

This is a cross-sectional longitudinal design that measures the differences between classrooms on quality and child outcomes. Students are nested in classrooms; classrooms are distinguished by teacher quality scores. Cross-sectional design was employed because the allocation of classroom quality relies on existing differences within the intervention. Longitudinal design was employed because the data does not fit within all cross-sectional design features including no time dimension and reliance on existing differences instead of change (De Vaus, 2001). The longitudinal design uses two measurements, pre-test in fall and post-test in spring.

Operationalization of Variables

The independent variable, teacher-child interactions, is measured by Classroom Assessment Scoring System CLASS, an observational assessment. Observers score based on three twenty minute observation cycles usually on the same day. CLASS is scored on a scale of 0-7 for each of the three areas of quality. The CLASS data is collected by PSESD coordinators in fall and spring for each classroom.

CLASS measures three domains: emotional support, instructional support and classroom organization on ten dimensions. Under emotional support the dimensions are positive climate, negative climate, teacher sensitivity and regard for students perspectives. The dimensions under instructional support are concept development, language modeling and quality of feedback. Classroom organization dimensions are behavior management, productivity and instructional learning formats. This study uses the three CLASS domains (emotional support, instructional support, and classroom organization) as independent variables rather than the ten underlying dimensions.

The dependent variable is child outcomes. This is measured by Teaching Strategies GOLD, an observational assessment that measures children's development in six areas, cognitive, language, literacy, mathematics, social-emotional, and physical development. Teachers rate the children's development on each objective 0-9. Examples of indicating behaviors for each level are provided for the raters by the GOLD instrument. Checkpoints are scored by teachers in the fall, winter and spring. This child assessment and planning

tool is currently being implemented in all Washington State kindergarten classrooms, a program called WAKids (SB5427, 2011-12).

This study also measures child outcome scores a second way, to account for the color bands representing developmentally appropriate age ranges for each objective score. As shown on figure 1, a score of five is developmentally appropriate for a prekindergarten four year old child (blue band) on objective 1. As shown on figure 2, a score of three is developmentally appropriate for a prekindergarten four year old child (blue band) on objective 19. Thus, GOLD scores have different meanings across learning objectives. GOLD assigns proficiency scores by age ranges Birth-1, 1-2 years, 2-3 years, Preschool 3, Pre-K 4 and Kindergarten. This study's proficiency scores place children in an age range based on the minimum sum of item scores to reach proficiency within a developmental area. This method of measuring student outcomes in proficiency bands results in an ordinal measure.

Figure 1 Teaching Strategies GOLD color bands

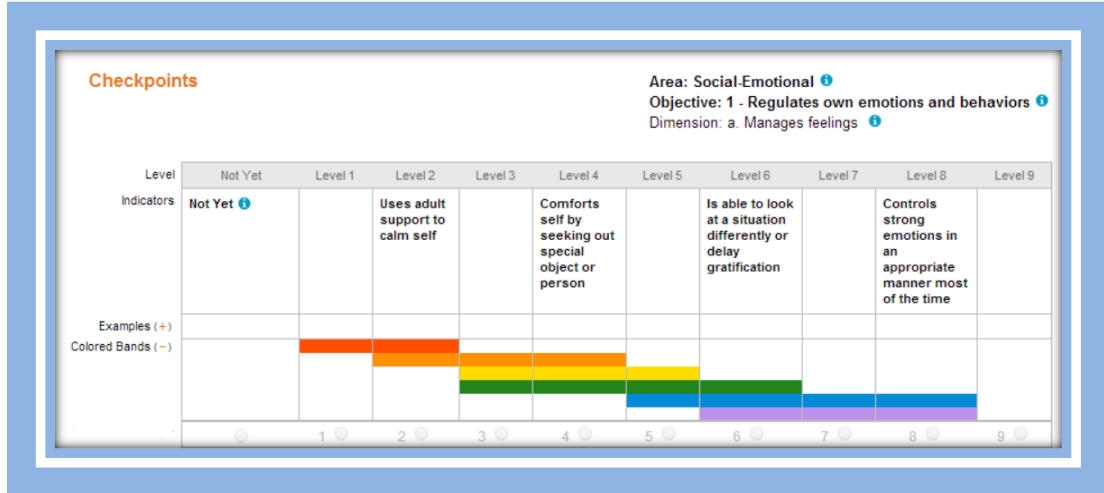
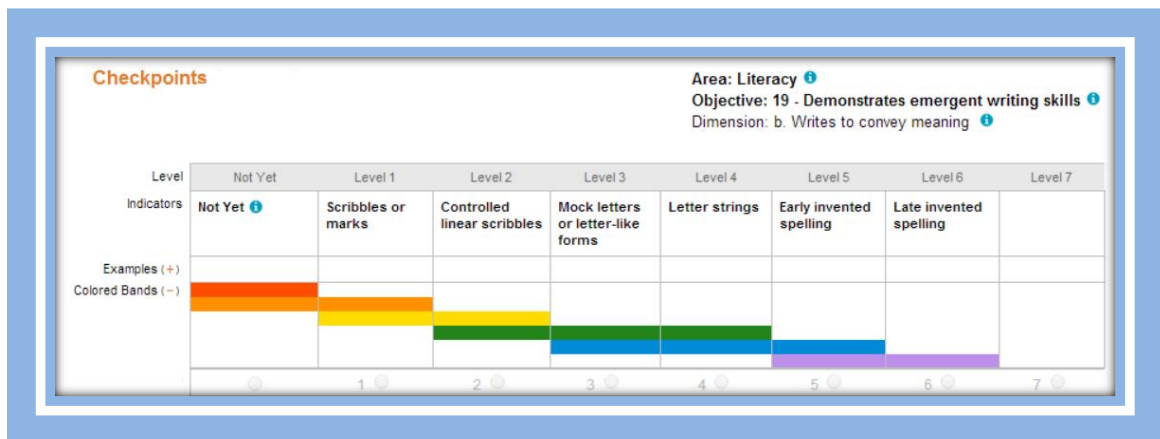


Figure 2 Teaching Strategies GOLD color bands



Validity

CLASS was found valid and reliable in a development and field testing research study using a nationally representative sample of 224 pre-kindergarten classrooms across 6 states (La Paro, 2004). Teaching Strategies assessment’s validity was tested using differential item functioning to compare the objective scores for children with and without diagnosed developmental delays and the checkpoints for English language learners and non-English language learners finding the tool to be equally reliable across all groups of children (Lambert, R. G., 2013).

This study's non-random sample using one educational service district creates an issue with external validity. The students were not randomly assigned to classrooms due to ethical and logistical reasons and the data collection had already occurred. The findings cannot directly generalize to other populations nor provide evidence that the results apply more widely. There are similarities between Head Start programs in PSESD to national Head Start preschools following federally mandated quality and program standards serving low income populations. PSESD evaluates teaching quality using the same assessment tool that are being adopted in Head Start programs nationally.

Data Analysis Methods

This study will use quantitative data at the student and classroom level of analysis. Descriptive statistics were used to present average child scores and CLASS quality scores and describe the distribution of the scores. Missing scores were deleted list wise. Two types of regressions were employed to analyze the relationship between teacher CLASS scores and child outcomes. Multilevel mixed effects regression nests children by classroom, a statistical model widely used in educational research. The multilevel mixed effects regression used children's average score across objectives within each developmental area. Multilevel regression was selected rather than classic ordinary least squared (OLS) regression because rather than obtaining coefficients by grouping units using a series of dummy variables, multilevel regression groups units and coefficients weighted based on reliability of the effect measurements for each group (Garson, 2013). A regression is run for each area of development, cognitive, language, literacy, mathematics, social-emotional, and physical development to discuss any developmental areas.

Ordinal logistical regression was used to analyze the CLASS proficiency band scores, ordered by age group. This method accounts for the variance in developmentally appropriate scores (color bands) that the raw scores do not account for.

Both regression models include an auto-regression of individual students fall scores to measure correlation of quality to child outcomes while accounting for student's baseline development. Independent variables in both regression models include race, language,

gender, Individual Education Plan (IEP, for children receiving services for developmental delays), spring age grade and the three classroom quality scores. These analytical methods will identify if quality has a relationship with child outcomes within specific areas of development. The expected results are children in classrooms with teachers scoring higher on quality of interaction will experience greater increases in developmental scores over one academic year in PSESD targeted prekindergarten programs.

Chapter 4- Results and Discussion

Children’s Developmental Scores

Factor analysis demonstrated the objectives within each of the six developmental areas are measuring the same latent variable (see appendix B for factor analysis tables). This study uses children’s average score for each developmental area.

The developmental scores range 0-9. The fall score means range from 2.8-5.7 with standard deviations of 1.1-1.3 for each developmental area. The spring score means range from 4.5-7 with standard deviations of 1-1.5 across developmental areas. The difference between fall and spring mean scores ranges from 1.3-1.7, the greatest mean gains in literacy and math (see table 1 below).

Table 1 Average Student Scores

Development Domain	Fall Mean	Fall standard deviation	Spring Mean	Fall standard deviation	Fall and Spring difference in means
Social Emotional	4.8	1.3	6.3	1.3	1.5
Physical	5.7	1.1	7	1	1.3
Language	5	1.3	6.4	1.3	1.4
Cognitive	4.6	1.1	6	1.2	1.4
Literacy	2.8	1.2	4.5	1.5	1.7
Math	3.5	1.3	5.2	1.5	1.7

Teacher-child interaction quality scores

Factor analysis demonstrated a high correlation of underlying dimensions within each domain, providing evidence that the dimensions are measuring the same underlying latent variables within each domain (see appendix A for CLASS factor analysis tables).

CLASS scores are between 0-7. The mean instructional support score is 4.7 with standard deviation of 1.17. The distribution of the instructional support scores is normal bell curve (see figure 3 below). Classroom organization score mean is 5.7 with a standard deviation of .84, the scores have a bell curve with a slight left skew (see figure 4 below). Emotional support score mean is 6 with a .84 standard deviation, the shape is left skewed (see figure 5 below). These averages are consistent with national CLASS score patterns (see table 2 below).

Table 2 Teacher CLASS score averages

Teacher CLASS Scores	Emotional Support	Classroom Organization	Instructional Support
FY 2013 National Average	5.99	5.63	2.72
PSESD Average	6	5.7	4.7

(Administration for Children & Families Early Childhood Learning & Knowledge Center, 2014).

Figure 3 CLASS scores Classroom Instructional Support

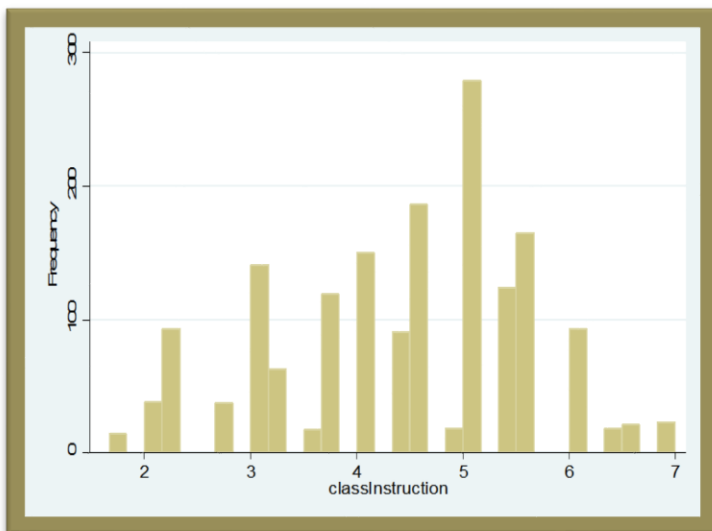


Figure 4 CLASS scores Classroom Organization

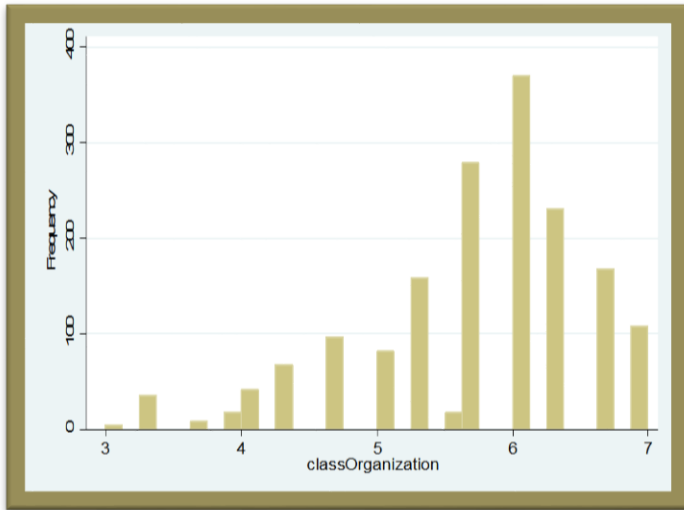
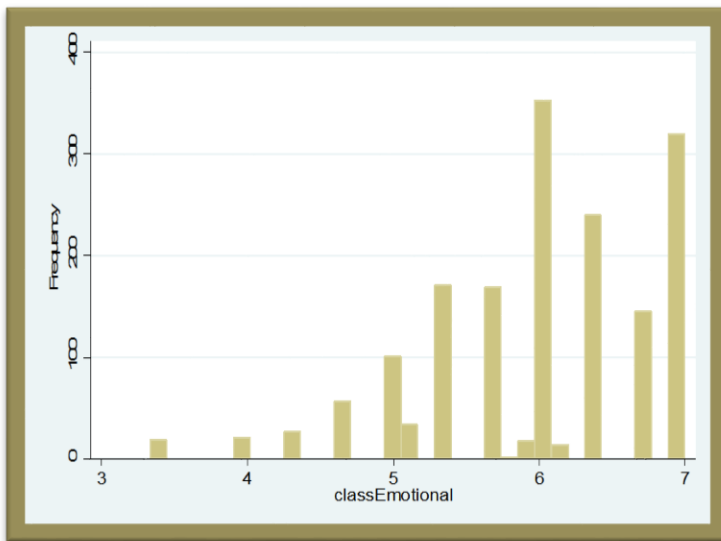


Figure 5 CLASS scores Classroom Emotional Support



Multilevel mixed effects regression

Below, table 3 shows the relationships between children’s GOLD spring scores and teachers’ CLASS scores, the previous GOLD score, and selected student characteristics using the multilevel mixed effects regression model. The results include a few positive relationships between classroom quality and child outcomes and several negative associations. There is a weak positive relationship between classroom organizational support ($b=.13, p=.08$) and physical development scores. Classroom instruction is weakly positively related to ($b=.13, p=.1$) cognitive development. Between

classroom emotional support and all developmental domains there is a low negative relationship ranging from $b=-.16$ to $b=-.22$ ($p \leq .1$).

Age is the accounts for the greatest amounts of variation of spring outcomes ($b=.57$ to $b=1.88$, $p=.00$). Fall score are highly related to the spring child development scores ranging from $b=.51$ to $b=.79$ ($p=.00$). Developmental delay has significant low to moderate negative relationship to student scores. Male had low negative relationships on social emotional, language and cognitive development. Race was not significantly related with development scores in any area. Language was not significantly related student scores except for a few languages in a few areas as shown on table 4 below. These language categories represented such a low percent (less than 5%) of the sample population the results can be skewed by a few individual children (see appendix c for full statistical outputs).

Table 3 Multilevel mixed effects regression

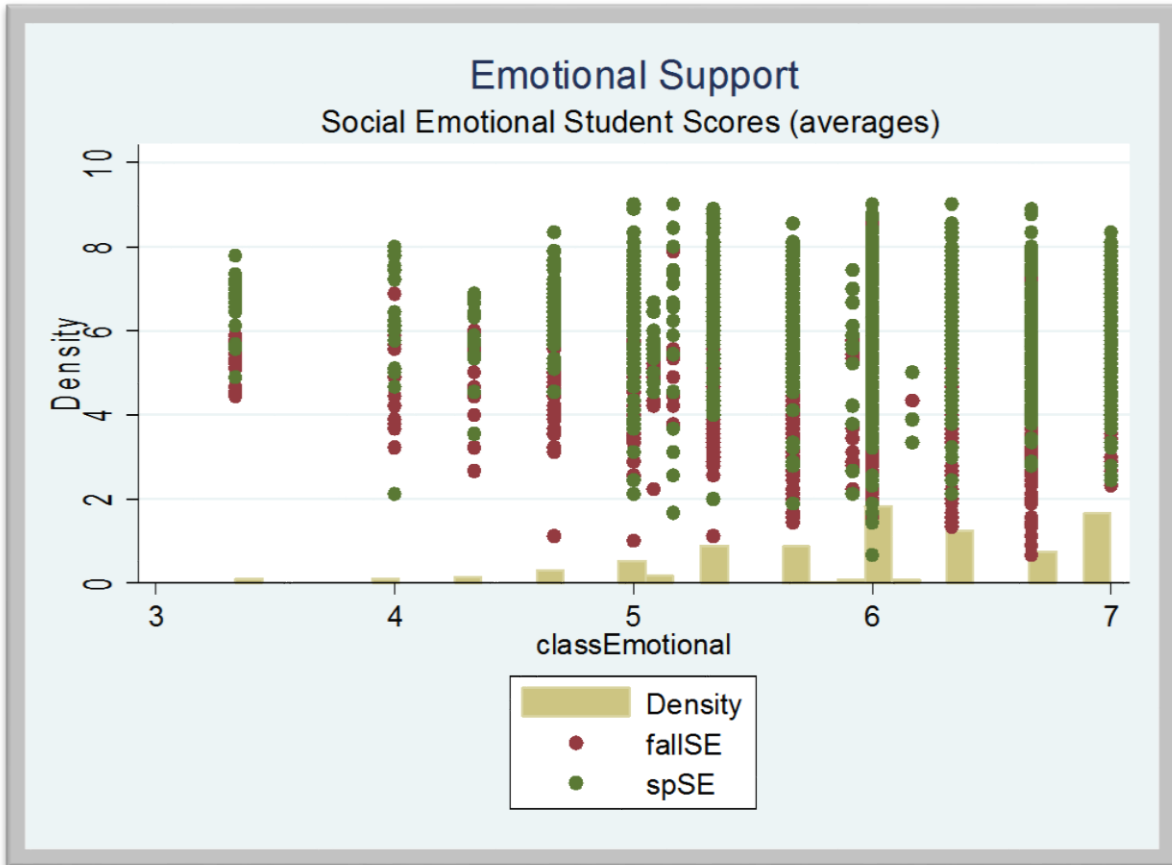
Classroom Teacher-Child Interaction	Social Emotional	Physical	Language	Literacy	Cognitive	Math
Organization	*	.13 (p=.08)	*	*	*	*
Instruction	*	*	*	*	.13 (p=.1)	*
Emotional Support	-.16 (p=.1)	-.17 (p=.02)	-.2 (p=.02)	-.22 (p=.04)	-.22 (p=.02)	-.16 (p=.1)
Age Group Preschool 3	.97 (p=.00)	.57 (p=.00)	.87 (p=.00)	.93 (p=.00)	1.02 (p=.00)	1.2 (p=.00)
Age Group Pre-K 4	1.5 (p=.00)	1.15 (p=.00)	1.32 (p=.00)	1.5 (p=.00)	1.52 (p=.00)	1.88 (p=.00)
Gender (male)	-.14 (p=.00)	*	-.06 p(.03)	*	-.07 (p=.01)	*
Fall Score	.57 (p=.00)	.51 (p=.00)	.63 (p=.00)	.79 (p=.00)	.66 (p=.00)	.68 (p=.00)
Developmental Delay	-.24 (p=.00)	-.18 (p=.00)	-.37 (p=.00)	-.36 (p=.00)	-.24 (p=.00)	-.37 (p=.00)

*Not statistically significant at the p<.1

Table 4 Language Mixed Regression (showing results significant at p=.05)

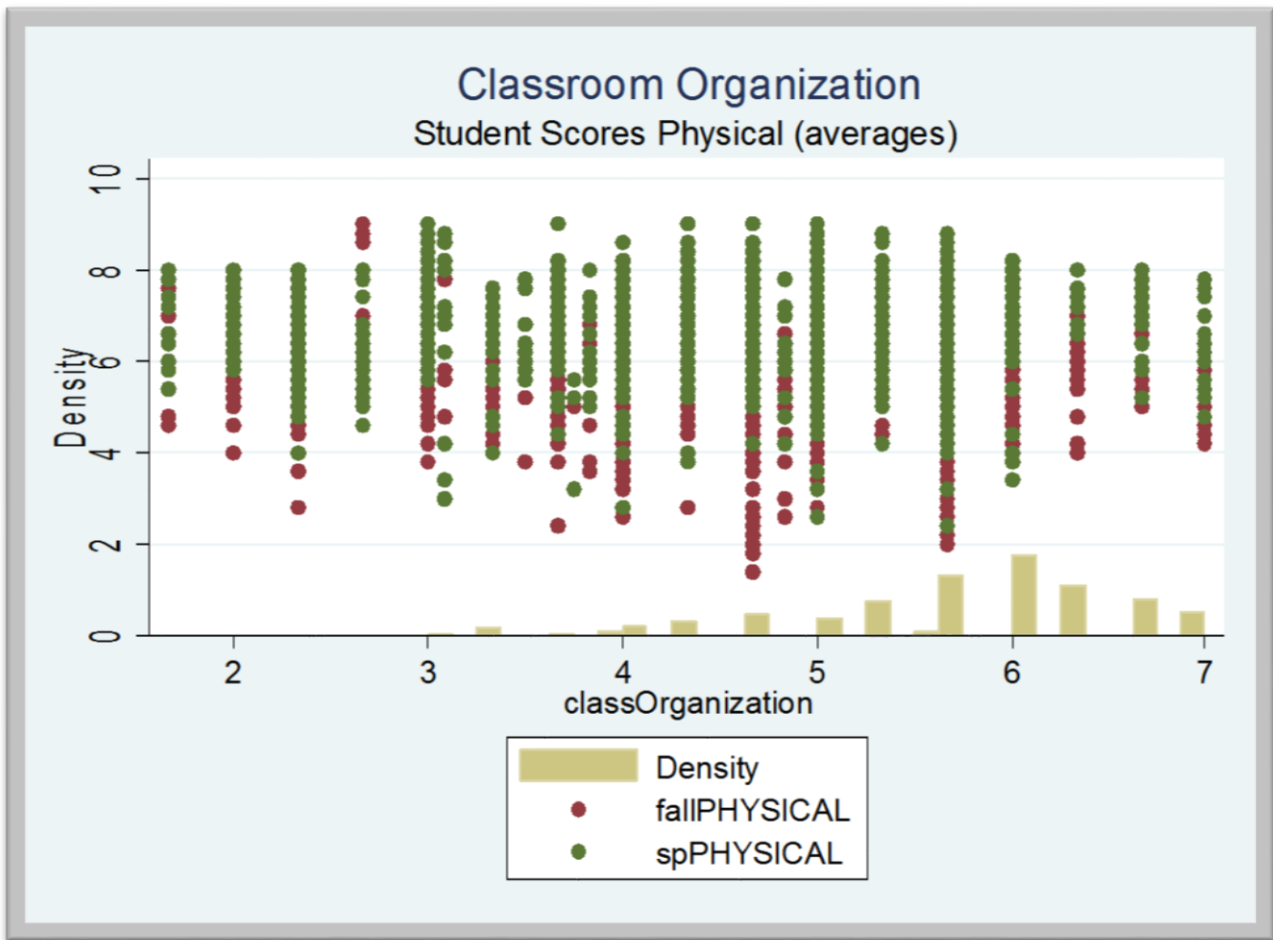
Language	Physical	Language	Literacy	Cognitive	Math
Punjabi	-.4 (p=.04)				
Chinese	.47 (p=.04)				
Arabic		-.33 (p=.05)		-.37 (p=.02)	
Ukrainian			-.52 (p=.03)		-.5 (p=.03)

Figure 6 Student Social Emotional Scores and Classroom Emotional Support



Shown in the figure 6 above, the classroom emotional support in the bars and the student social emotional scores in the maroon (fall) and green dots (spring). Classroom emotional support is negatively related to social emotional development ($b=-.16, p=.1$). This graph shows the green dot student spring scores have a slight downward trend as the classroom emotional support scores rise to the 6-7 range, although the relationship is weak.

Figure 7 Student physical development scores and classroom organization scores



The figure 7 above shows the classroom organization scores in the bars and student physical development scores in the marron(fall) and green dots (spring). There is a weak positive relationship between class organization and physical development ($b=.13, p=.08$). Although it is difficult to identify a positive relationship on figure 7 it does not have the same downward trend as seen on the previous figure 6.

Ordinal Logistical Regression

Utilizing the age proficiency bands as dependent variables and ordinal logistical regression, the results are similar to the mixed effects model. These coefficients show for every one unit increase in classroom organizational quality children’s scores are expected

to increase by .48 units ($p=.00$) in physical development and .19 units ($p=.1$) in language development. For every unit increase in classroom instruction quality, student scores are predicted to increase .25 ($p=.002$) in social emotional and .13 ($p=.1$) in cognitive development. Again emotional support quality predicts lower students scores. For each increase in emotional support, student scores are expected to decrease by -.43 in social emotional, -.48 in physical, -.46 in language, -.4 in literacy, -.49 in cognitive and -.2 in math development ($p\leq.07$).

Developmental delay predicted lower scores between -.44 and -.9 units ($p\leq.01$). For a unit increase in fall score, spring scores are expected to increase .1 to .18 ($p=.00$). Age is again the best predictor of spring scores, preschool age 3 group is predicted to have 2.3 to 2.54 ($p\leq.01$) increase in spring score. Pre-K age 4 group is expected to have increases 1.2 to 4.33 ($p\leq.03$) (see table 5). Race is not related to student scores except for a few areas shown on table 7 below. Language is not significantly related except for a few languages and areas shown in table 6 below (for full statistical outputs see Appendix D).

Table 5 Ordinal logistical regression

Classroom Teacher-Child Interaction Quality Score	Social Emotional	Physical	Language	Literacy	Cognitive	Math
Organization	*	.48 (p=.00)	.19 (p=.1)	*	*	*
Instruction	.25 (p=.002)	*	*	*	.13 (p=.1)	*
Emotional Support	-.43 (p=.001)	-.48 (p=.001)	-.46 (p=.00)	-.4 (p=.00)	-.49 (p=.00)	-.2 (p=.07)
Age Group Preschool 3	1.54 (p=.001)	5.3 (p=.00)	2.3 (p=.00)	1.8 (p=.005)	2.54 (p=.00)	2.16 (p=.00)
Age Group Pre-K 4	3.33 (p=.00)	1.2 (p=.03)	4 (p=.00)	3.5 (p=.00)	4.33 (p=.00)	3.76 (p=.00)
Gender (male)	-.55 (p=.00)	-.26 (p=.05)	-.32 p(.01)	*	-.26 (p=.06)	*
Fall Score	.1 (p=.00)	.17 (p=00)	.12 (p=.00)	.14 (p=.00)	.13 (p=.00)	.18 (p=.00)
Developmental Delay	*	-.53 (p=.01)	-.9 (p=.00)	-.78 (p=.00)	-.6 (p=.01)	-.44 (p=.00)

*Not statistically significant at the p<.1

Table 6 Language Ordinal Regression (showing results significant at the p=.05)

Language	Physical	Language	Math
Spanish	.83 (p=.03)		
Russian		-1.58 (p=.01)	
Arabic			
Ukrainian			-1.62 (p=.01)

Table 7 Race Ordinal Regression (showing results significant at p=.05)

Race	Social Emotional	Language	Literacy	Cognitive
Native Hawaiian and Pacific Islander	1.33 (p=.03)			1.78 (p=.01)
Asian		.84 (p=.05)		

Discussion

The Null Hypothesis, no relationship between teachers' quality and student outcomes, cannot be rejected because there is not sufficient strength in coefficients. However the significant negative coefficients across all developmental domains for CLASS emotional support is unusual when compared to other research using CLASS and other teacher emotional support measures. This raises questions around how CLASS is being scored in classroom setting. Such as could rater bias, the effect on teachers being observed or the frequency and length of CLASS observations impact the validity of the scores? This study suggests subjectivity and the systems around CLASS scoring in early childhood setting should be subject to future research.

This study finds that teachers' classroom organization quality is related to children's physical outcomes. It was expected that classroom productivity and behavior management would contribute to children's opportunities for fine and gross motor practice. If the children are engaged and managing transitions efficiently there is more time for active and hands on activities indoors and outdoors.

Similarly this study's finding that teachers' classroom instruction is related to children's social emotional and cognitive development is consistent with the literature (Mashburn, A. J., 2008). The relationship of age and fall score to child outcomes are logical and expected. Teaching strategies GOLD scoring scale is developed out of literature documenting typical child developmental patterns by age. The fall score is a child's baseline score, it is expected the spring score will be relatively higher than the fall score. For example, a child scoring low in the fall is less likely to score high in spring than a child who had already scored high in the fall.

The surprising finding is the negative relationship between teachers' classroom emotional support and children's social, emotional, cognitive, and physical outcomes. The relationship is weak however there is a consistent pattern of this negative relationship across every developmental area. There is no evidence in the literature that supporting young children emotionally in classrooms results in decreased developmental gains. As referenced earlier many studies find positive correlations between classroom emotional support and children's development using CLASS (Mashburn, A. J., 2008, Moiduddin, et al., 2012). This leads to questioning of the process in which classroom emotional support is measured by the CLASS raters who visit classroom.

Another aspect of the class emotional support is the high national and PSESD averages. Out of a scale of 7 the average in this sample is 6 and the national average 5.99, which seems very high. The length of observation is only 40-60 minutes, only one day. It would be difficult to adjust the productivity of the children, efficiency of transitions, instruction, and quality of feedback and behavior management for just one day. However during one hour of observation the effect of being observed may change a teacher's level of friendliness, warmth, teacher sensitivity, regard for student perspective and the score may not accurately represent the classroom emotional support that the children experience during the entire school year.

Finally, there is a possible correlation among students who have higher social emotional needs performing lower on the developmental scores while their teachers, responding to student emotional needs, score higher on emotional support. One report

found high teacher-child interaction quality has higher impacts on student's social competence and lower problem behavior when present in classrooms with higher overall quality (Burchinal, 2010). This indicates for classroom emotional support to be effective the classroom requires overall quality. The structural classroom quality measures such as the building or class size that are not included in this data may work in conjunction with emotional support impacting child outcomes.

Chapter 5- Conclusions

In the multilevel mixed effects regression teachers' instructional and organizational quality had positive relationships to physical and cognitive development. In the logistical ordinal regression teachers' instructional and organizational quality had positive relationships to children's physical, cognitive, language and social-emotional development. Positive low-moderate relationships are consistent with previous literature relating quality to outcomes. In these areas this study's results are modest yet predicted and supported by previous research.

Teachers' emotional support quality had a negative relationship to children's developmental outcomes in each of the six areas, physical, language, literacy, cognitive, math and social-emotional development. These findings are counter to previous literature relating teachers' emotional support quality to child outcomes. While the relationship is weak the consistency across developmental domains indicates there may be a validity issue in how CLASS measures emotional support. Threats to validity of the CLASS emotional support score include teachers' displaying unusual levels of emotional support in reaction to being observed and rater bias. The effect of being observed on teachers' behavior could be an issue of how frequently and for what duration teachers are observed. Further research is recommended to develop best practices in the use of CLASS.

The Office of Head Start outlined the use of CLASS is a 2012 policy brief; CLASS observations can be focused on at least three different levels:

- “1. for **monitoring purposes**, such as the triennial review process conducted by the Office of Head Start;
2. for **program planning and evaluation**, such as when a grantee conducts their own classroom observations to evaluate quality or plan professional development; and
3. for **assessment of individual classrooms**, as part of individualized professional development planning.” (Office of Head Start, 2012).

Federally funded programs use CLASS as a high stakes assessment to manage grantee designations. State quality rating and improvement systems, including Washington's Early Achievers, offer higher reimbursements for higher quality scores. Many States are utilizing CLASS as part of the quality assessment. However based on the current study a few significant questions should be addressed before CLASS is utilized to shape the professional development and rating for early learning programs.

Does this tool function as a better quality measure when used with a set of quality measures including structural measures? Some research indicates teachers' quality has greater relationship to student outcomes in overall higher quality classrooms (Burchinal, 2010). CLASS may serve as a better indicator of teachers' quality when the scores are used within the larger context of the classroom. Such as how big is the school, percentage of students with developmental delays, teacher-child ratio, teachers' access to school supplies and physical space.

How can inter-rater reliability be addressed and how often should data be collected? To identify possible rater bias classrooms could be observed by multiple raters and the scores compared or averaged. Scores might have higher validity if the classroom was visited every month, or randomly, rather than a scheduled yearly observation. Further research on best practices using CLASS could identify optimal length of observation to accurately sample teacher-child interaction.

Suggested ways to further validate CLASS include comparing CLASS scores to other measures of teacher-interaction quality and independent studies of CLASS validity to avoid bias. Many of the studies validating CLASS are conducted with Robert Pianta, the developer of the assessment (La Paro, 2004). There is an assumption growing in the field that widespread use of CLASS is equivalent to validity. "CLASS Pre-K is an observational tool based on more than a decade of research on teacher-child interactions in more than 10,000 classrooms across a wide range of communities and programs" (Office of Head Start, 2012). Most of these studies utilized CLASS but were not for the purpose of studying validity of CLASS or the implementation of CLASS. Research exploring best practices in how to use

CLASS for greatest validity will further the usefulness of CLASS in professional development, research and policy.

There are several limitations in the current study. The CLASS scores taken in the fall measure the classroom quality regardless of teacher turnover, from this data there is no way to tell if the classroom quality in the fall is an accurate measure for the whole academic year. The student outcome scores on Teaching Strategies Gold are teacher rated and subjective, there could be inter-rater reliability issues.

Teaching strategies GOLD checkpoints are valuable in the wide range of developmental areas covered and specific items that present a holistic image of children. The benefit to being teacher rated is children are observed in their natural setting without testing disruption. This can be more accurate than testing children because their knowledge is contextual and they lack motivation, attention, and cognitive skills for many types of testing. However for research, the Teaching Strategies GOLD data presents several challenges. The raw scores are not comparable, they don't reflect a developmentally appropriate skill level because the scores have different meanings across objectives. Setting age proficiency bands is cumbersome because ordinal data limits statistical models appropriate. The range is low (0-9) and with this low range it is difficult to observe variance between children across classrooms. The standard deviations from the means in fall and spring scores was always lower than 1.5.

In addition to accountability these observational tools impact educational settings for staff and teachers, curriculum and resources. If the field is to focus on a few set of observational tools measuring quality and children's development those tools should reflect the goals and values in education and have clearly articulated purpose in addition to being reliable and valid. In this study CLASS instructional support and organization scores and child outcomes demonstrate the possibility of quality improvement to relate to child outcomes. Further development on the practical use of CLASS rating system will better validate the social emotional quality score and bolster greater confidence in the use of CLASS for professional development and decision making.

References

- Ackerman, Debra J. (2014). *State Funded PreK Policies on External Classroom Observation*. Policy evaluation and research center, Educational testing service. Retrieved from <http://www.ets.org/Media/Research/pdf/PIC-STATE-PRE-K.pdf>
- Administration for Children & Families Early Childhood Learning & Knowledge Center. (May, 2014). Use of Classroom Assessment Scoring System in Head Start. Retrieved from <http://eclkc.ohs.acf.hhs.gov/hslc/hs/sr/class>.
- Brophy-herb, H. E., Lee, R. E., Nievar, A., & Stollak, G. (2007). Preschoolers' social competence: Relations to family characteristics, teacher behaviors and classroom climate. *Journal of applied developmental psychology* 28, 134-148.
- Burchinal, M., Kainz, K., & Cai, Y. (2011). How Well do our Measures of Quality Predict Child Outcomes? A meta-analysis and coordinated analysis of data from large-scale studies of early childhood settings. In Zaslow, M., Martinez-beck, I., Tout, K., Halle, T. (Eds.), *Quality Measurement in Early Childhood Settings* (11-31). Baltimore, MD: Paul H Brookes Publishing.
- Burchinal, M., Vandergrift, N., Pianta, R., & Mashburn, A. (2010). Threshold analysis of association between child care quality and child outcomes for low-income children in pre-kindergarten programs. *Early Childhood Research Quarterly* 25, 166-176.
- De Vaus, David. (2001). *Research Design in Social Research*. Thousand Oaks: SAGE Publications, Inc.
- Downer, J. T., Lopez, M. L., Grimm, K. J., Hamagami, A., Pianta, R. C., & Howes, C. (2012). Observations of teacher-child interactions in classrooms serving Latinos and dual language learners: Applicability of the Classroom Assessment Scoring System in diverse settings. *Early Childhood Research Quarterly* 27, 21-32.
- Downer, J., Sabol, T., Hamre, B. (2010). Teacher-Child Interactions in the Classroom: Toward a Theory of within and Cross Domain Links to Children's Developmental Outcomes.

- Driscoll, K. C., Wang, L., Mashburn, A. J., & Pianta, R. C. (2011). Fostering Supportive Teacher-Child Relationships: Intervention in a State-Funded Preschool Program. *Early Education and Development* 22(4), 593-619.
- Early, D., Barbarin, O., Bryant, D., Burchinal, M., Chang, F., Clifford, R., Crawford, G., ... Barnett, S. W. (2005). Pre-Kindergarten in Eleven States: NCEDL's Multi-State Study of Pre-Kindergarten & Study of State-Wide Early Education Programs (SWEEP). NCEDL Working Paper. Retrieved from <http://fcd-us.org/sites/default/files/Prekindergartenin11States.pdf> (10/22/13).
- Early, D. M., Bryant, D. M., Pianta, R. C., Clifford, R. M., Burchinal, M. R., Ritchie, S., Howes, C., & Barbarin, O. (2006). Teachers' education, major, and credentials related to classroom quality and children's academic gains in prekindergarten? *Early Childhood Research Quarterly* 21(2), 174-195.
- Early, D. M., et al. (2007) Teachers' education, Classroom Quality, and Young Children's Academic Skills: Results from Seven Studies of Preschool Programs. *Child Development* 78(2), 558-580.
- Early Childhood Learning and Knowledge Center (ECLKC). U.S. Department for Health and Human Services, Administration for Children and Families. Head Start Statutory Degree and Credentialing Requirements for Head Start Teaching Staff. Retrieved from http://eclkc.ohs.acf.hhs.gov/hslc/standards/IMs/2008/resource_012_0081908.html. (10/22/14).
- Early Childhood Learning and Knowledge Center (ECLKC). U.S. Department for Health and Human Services, Administration for Children and Families. Report to Congress The Final Head Start Program Designation Renewal System. Retrieved from http://eclkc.ohs.acf.hhs.gov/hslc/mr/rc/Head_Start_Designation_Renewal_System_Final_Rule.pdf (10/22/14).

- Garson, David G. (2013) Chapter 1. Fundamentals of Hierarchical Linear and Multilevel Modeling. Garson, David G. (ed.), *Hierarchical Linear Modeling: guide and applications*. (3-26). California: SAGE Publications, Inc.
- Googe, H. S. (2011). *A Short-Term Longitudinal Study of the Relationship Between Classroom Quality and Child Language and Academic Outcomes in a State-Funded Pre-Kindergarten Program*. (Unpublished doctoral dissertation). University of South Carolina.
- Gosse, C. S., McGinty, A. S., & Mashburn, A. J. (2012). The Role of Relational and Instructional Classroom Supports in the Language Development of At-Risk Preschoolers. *Society for Research on Educational Effectiveness (SPREE)*. Spring Conference Abstract.
- Hamre, B. K., Pianta, R.C., Burchinal, M., Field, S., LoCasale-Crouch, J., Downer, J. T., Howes, C., et al. (2012). A Course on Effective Teacher-Child Interactions: Effects on Teacher Beliefs, Knowledge, and Observed Practice. *American Educational Research Journal* 49(1), 88-123.
- Helburn, S. W., Ed. (1995). Cost, Quality and Child Outcomes in Child Care Centers. Key Findings and Recommendations. *Young Children* 50(4), 40-44.
- Howes, C. (1997). Children's experiences in center-based care as a function of teacher background and adult: child ratio. *Merrill-Palmer Quarterly* 43(3), 404-425.
- Howes, C., Burchinal, M., Pianta, R., Bryant, D., Early, D., Clifford, R., & Barbarin, O. (2008) Ready to Learn? Children's pre-academic achievement in pre-kindergarten programs. *Early Childhood Research Quarterly* 23, 27-50.
- Keys, T. D., Farkas, G., Burchinal, M. R., Duncan, G. J., Vandell, D. L., Li, W., Ruzek, E. A., & Howes, C. (2013). Preschool Center Quality and School Readiness: Quality Effects and Variation by Demographic and Child Characteristics. *Child Development* 84(4), 1171-1190.

- La Paro, K. M., Pianta, R. C., & Stuhlman, M. (2004). The Classroom Assessment Scoring System: Findings from the Prekindergarten Year. *The Elementary School Journal* 104(5), 409-426.
- Lambert, R. G., Kim, D. H., & Burts, D. C. (2013). Using teacher ratings to track the growth and development of young children using teaching strategies GOLD assessment system. *Journal of Psychoeducational Assessment*, XX(X), 1-13
- Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D., Burchinal, M., Early, D. M., Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language and social skills. *Child Development* 79(3), 732-749.
- Moiduddin, E., Aikens, N., Tarullo, L., West, J., Xue, Y. (2012). *Child Outcomes and Classroom Quality in FACES 2009. OPRE Report 2012-37a*. Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. Washington, DC.
- NICHD Early Childhood Research Network. (2006). Child care effect size for the NICHD Early Child Care and Youth Development. *American Psychologist* 61(2), 99-116
- NICHD Early Child Care Research Network. (2005) Early Child Care and Children's Development in the Primary Grades: Follow-Up Results from the NICHD Study of Early Child Care. *American Educational Research Journal* 42(3), 537-570.
- Office of Head Start National Center on Teaching Quality and Learning. (2012). *Understanding and Using CLASS for Program Improvement*. Retrieved from <http://eclkc.ohs.acf.hhs.gov/hslc/tta-system/teaching/docs/class-brief.pdf>
- Pianta, R., Howes, C., Burchinal, M., Bryant, D., Clifford, R., Early, D., & Barbarin, O. (2005). Features of Pre-Kindergarten Programs, Classrooms and Teachers. Do They Predict Observed Classroom Quality and Child-Teacher Interactions? *Applied Developmental Science* 9(3), 144-159.

- Puma, M., Bell, S., Cook, R., & Heid, C. (2010). *Head Start Impact Study*. U.S. Department of Human and Health Services, Administration for Children and Families. Washington, DC.
- Peisner-Feinberg, E. S., Burchinal, M. R., Clifford, R. M., Culkin, M. L., Howes, C., Kagan, S. L., & Yazejian, N. (2001) The Relation of Preschool Child-Care Quality to Children's Cognitive and Social Development Trajectories through Second Grade. *Child Development* 72(5), 1534-1553.
- Sabol, T. J., Soliday Hong, S. L., Pianta, R. C., & Burchinal, M. R. (2013). Can Rating Pre-K Programs Predict Children's Learning? *Science* 341, 845-846.
- SB5427, 2011-12. retrieved from;
<http://apps.leg.wa.gov/billinfo/summary.aspx?bill=5427&year=2011>
- Vi-Nhuan, L., Kirby, S. N., Barney, H., Setodji, M. C., Gershwin, D. (2006). *School Readiness, Full-Day Kindergarten, and Student Achievement: An Empirical Investigation*. Santa Monica, CA: RAND Corporation.
- Wasik, B. A., & Hindman, A. H. (2011). Improving Vocabulary and Pre-Literacy Skills of At-Risk Preschoolers Through Teacher Professional Development. *Journal of Educational Psychology* 103(2), 455-469.
- Zaslow, M., Anderson, R., Redd, Z., Wessel, J., Tarullo, L. & Burchinal, M. (2010). *Quality Dosage, Thresholds, and Features in Early Childhood Settings: A Review of the Literature, OPRE 2011-5*. Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. Washington, DC.
- (2007). *School Readiness Survey*. U.S. Department of Education, National Center for Educational Statistics, National Household Educational Survey (NHES).

Appendixes Table of Contents

- A. Factor analysis CLASS domains and dimensions
- B. Factor analysis Teaching Strategies Gold developmental domains and objectives
- C. Multilevel Mixed Effects Regressions
 - 1. Social Emotional
 - 2. Physical
 - 3. Language
 - 4. Literacy
 - 5. Cognitive
 - 6. Mathematics
- D. Ordinal Logistical Regressions
 - 1. Social Emotional
 - 2. Physical
 - 3. Language
 - 4. Literacy
 - 5. Cognitive
 - 6. Mathematics
- E. Population Description
 - 1. Language Table and Bar Graph
 - 2. Race Table and Pie Chart
 - 3. Age Group Table

Appendix A Factor analysis CLASS domains and dimensions

A.1 Emotional Support- Positive Climate, Teacher Sensitivity, Regard for Student Perspectives (excludes Negative Climate because of reverse scoring)

Variable	Factor1	Uniqueness
PositiveCl~e	0.7153	0.4883
TeacherSen~y	0.8046	0.3527
StudentPersp	0.7710	0.4056

A.2 Classroom Organization- Behavior Management, Productivity, Instructional Learning Formats

Variable	Factor1	Uniqueness
BehaviorMa~t	0.7787	0.3936
Productivity	0.8504	0.2769
InstrLearn~s	0.6935	0.5191

A.3 Classroom Instructional Support- Concept Development, Quality of Feedback and Language Modeling

Variable	Factor1	Uniqueness
ConceptDev	0.9261	0.1424
QualFeedback	0.9068	0.1777
LanguageMo~g	0.8250	0.3193

Appendix B Factor analysis Teaching Strategies Gold developmental domains and objectives

B.1 Social Emotional

Variable	Factor1	Factor2	Factor3	Factor4	Uniqueness
fallSocia~1a	0.7934	0.1699	-0.0958	-0.0327	0.3314
fallSocia~1b	0.8107	0.1918	-0.1385	-0.0029	0.2868
fallSocia~1c	0.8123	0.2398	0.0114	0.1001	0.2725
fallSocia~2a	0.7566	0.2082	0.1480	0.0703	0.3573
fallSocia~2b	0.8399	-0.0585	0.1166	-0.0791	0.2712
fallSocia~2c	0.8051	-0.2804	-0.0353	0.0894	0.2640
fallSocia~2d	0.7803	-0.2883	0.0099	0.0952	0.2988
fallSocia~3a	0.8598	-0.1359	-0.0839	-0.0807	0.2288
fallSocia~3b	0.8322	-0.0312	0.0744	-0.1401	0.2813

Variable	Factor1	Factor2	Factor3	Factor4	Uniqueness
springSoc~1a	0.8408	0.0816	-0.1497	0.0184	0.2637
springSoc~1b	0.8624	0.1756	-0.1099	0.0461	0.2112
springSoc~1c	0.8447	0.2598	0.0642	0.0246	0.2143
springSoc~2a	0.7807	0.2096	0.1470	-0.0276	0.3242
springSoc~2b	0.8864	-0.0317	0.0157	-0.0888	0.2051
springSoc~2c	0.8820	-0.1846	0.0646	0.0620	0.1799
springSoc~2d	0.8523	-0.2052	0.0912	0.0524	0.2205
springSoc~3a	0.9064	-0.1550	-0.0712	0.0093	0.1493
springSoc~3b	0.8727	-0.1110	-0.0372	-0.0958	0.2156

B.2 Physical

Variable	Factor1	Factor2	Uniqueness
----------	---------	---------	------------

fallPhysic~4		0.8171	-0.1656		0.3050
fallPhysic~5		0.8595	-0.1409		0.2413
fallPhysic~6		0.7731	-0.1406		0.3825
fallPhysi~7a		0.7691	0.2335		0.3539
fallPhysi~7b		0.6659	0.2786		0.4789

Variable		Factor1	Factor2		Uniqueness
springPhys~4		0.8668	-0.2052		0.2065
springPhys~5		0.8806	-0.1932		0.1872
springPhys~6		0.8376	-0.1592		0.2731
springPhy~7a		0.8294	0.2810		0.2332
springPhy~7b		0.7814	0.3178		0.2884

B.3 Language

Variable		Factor1	Factor2	Factor3		Uniqueness
fallLangu~8a		0.8560	0.2283	-0.0314		0.2141
fallLangu~8b		0.8047	0.2540	0.0301		0.2871
fallLangu~9a		0.8938	0.0551	-0.0629		0.1942
fallLangu~9b		0.8400	-0.0684	-0.0498		0.2872
fallLangu~9c		0.8900	-0.1915	-0.0261		0.1705
fallLangu~9d		0.9038	-0.1235	-0.0154		0.1676
fallLang~10a		0.8781	-0.1297	0.0604		0.2084
fallLang~10b		0.8699	0.0052	0.0975		0.2337

Variable	Factor1	Factor2	Factor3	Uniqueness
springLan~8a	0.8774	0.1541	0.0496	0.2040
springLan~8b	0.8414	0.2000	-0.0035	0.2521
springLan~9a	0.8950	-0.0055	0.1063	0.1877
springLan~9b	0.8786	-0.0851	0.0898	0.2127
springLan~9c	0.9175	-0.1393	-0.0003	0.1388
springLan~9d	0.9198	-0.0911	-0.0123	0.1455
springLa~10a	0.9021	-0.0673	-0.1032	0.1710
springLa~10b	0.8843	0.0549	-0.1243	0.1995

B.4 Literacy

Variable	Factor1	Factor2	Factor3	Factor4	Factor5	Uniqueness
fallLite~15a	0.7434	-0.1564	0.1522	0.2005	-0.0023	0.3595
fallLite~15b	0.8463	-0.1563	0.2134	0.1126	0.0315	0.2001
fallLite~15c	0.7825	-0.1650	0.1157	0.0618	0.0683	0.3385
fallLite~16a	0.7110	0.1802	0.2320	-0.1367	-0.0768	0.3836
fallLite~16b	0.7845	0.0573	0.1740	-0.2295	-0.0318	0.2973
fallLite~17a	0.7707	-0.0659	-0.1226	-0.0908	0.0722	0.3731
fallLite~17b	0.8329	-0.0287	-0.1067	-0.1210	0.1306	0.2623
fallLite~18a	0.8828	-0.1129	-0.1478	0.0119	-0.0838	0.1790
fallLite~18b	0.8755	-0.1097	-0.1816	-0.0333	-0.0121	0.1872
fallLite~18c	0.8534	-0.1443	-0.1657	0.0445	-0.1148	0.2083
fallLite~19a	0.6507	0.4678	-0.0326	0.0991	0.0087	0.3469
fallLite~19b	0.6812	0.4312	-0.0741	0.1049	0.0199	0.3332

Variable	Factor1	Factor2	Factor3	Factor4	Factor5	Uniqueness
springLi~15a	0.8002	-0.1079	0.1140	0.1844	0.0294	0.3002

springLi~15b		0.8970	-0.0989	0.1845	0.0812	-0.0330		0.1439
springLi~15c		0.8507	-0.0848	0.1789	0.0887	-0.0844		0.2221
springLi~16a		0.7730	0.1305	0.1958	-0.0933	0.1061		0.3272
springLi~16b		0.8365	-0.0210	0.2044	-0.1838	0.0330		0.2231
springLi~17a		0.8423	-0.0562	-0.1266	-0.0826	-0.0521		0.2618
springLi~17b		0.8756	-0.0297	-0.0626	-0.1378	-0.0926		0.2009
springLi~18a		0.8917	-0.1443	-0.1591	0.0429	0.0192		0.1565
springLi~18b		0.8973	-0.1035	-0.1752	-0.0364	0.0110		0.1521
springLi~18c		0.8844	-0.1326	-0.1896	0.0384	0.0993		0.1531
springLi~19a		0.7270	0.4085	-0.0693	0.0822	0.0017		0.2931
springLi~19b		0.8095	0.3448	-0.0651	0.0260	-0.0282		0.2201

B.5 Cognitive

Variable		Factor1	Factor2	Factor3	Factor4		Uniqueness
fallCogn~11a		0.8124	-0.1989	0.1033	0.0211		0.2892
fallCogn~11b		0.8514	-0.2129	0.0659	-0.0029		0.2255
fallCogn~11c		0.8444	-0.1458	-0.0619	-0.0595		0.2583
fallCogn~11d		0.8605	-0.0916	-0.1046	0.0265		0.2395
fallCogn~11e		0.8551	-0.0083	-0.0836	0.0825		0.2549
fallCogn~12a		0.8392	0.1296	-0.0485	-0.0918		0.2681
fallCogn~12b		0.8645	0.0538	0.0099	-0.0482		0.2473
fallCogni~13		0.8194	0.1501	0.0676	-0.0387		0.3000
fallCogn~14a		0.8270	0.1857	0.0468	0.0298		0.2784
fallCogn~14b		0.8005	0.1500	0.0146	0.0852		0.3292

Variable		Factor1	Factor2	Factor3	Factor4		Uniqueness
springCo~11a		0.8759	-0.1754	0.0625	0.0167		0.1979

springCo~11b		0.8916	-0.1764	0.0358	0.0160		0.1724
springCo~11c		0.8979	-0.0824	-0.1018	0.0162		0.1764
springCo~11d		0.8987	-0.0381	-0.0883	-0.0434		0.1812
springCo~11e		0.9135	0.0140	-0.0637	-0.0515		0.1586
springCo~12a		0.8838	0.1663	-0.0614	0.0246		0.1869
springCo~12b		0.9071	0.0635	-0.0704	0.0421		0.1664
springCog~13		0.8879	0.1063	0.0729	0.0217		0.1946
springCo~14a		0.8947	0.0522	0.1235	-0.0020		0.1815
springCo~14b		0.8654	0.0702	0.0983	-0.0402		0.2348

B.6 Mathematics

Variable		Factor1	Factor2		Uniqueness
fallMath20a		0.8335	0.1590		0.2800
fallMath20b		0.8479	0.1517		0.2581
fallMath20c		0.7878	0.1902		0.3433
fallMath21a		0.8151	-0.0710		0.3306
fallMath21b		0.8384	-0.0716		0.2919
fallMath22		0.8565	-0.1884		0.2308
fallMath23		0.7967	-0.1652		0.3379

Variable		Factor1	Factor2		Uniqueness
springMa~20a		0.8571	0.1843		0.2315
springMa~20b		0.9086	0.1152		0.1612
springMa~20c		0.8706	0.1924		0.2051
springMa~21a		0.8657	-0.1546		0.2266
springMa~21b		0.8859	-0.0875		0.2075

```

springMath22 | 0.9007 -0.1502 | 0.1662
springMath23 | 0.8609 -0.0969 | 0.2495

```

Appendix C Multilevel Mixed effects regression

C.1 Social Emotional regression

```

Mixed-effects ML regression      Number of obs   =    1396
Group variable: ClassID         Number of groups =    110

                                Obs per group: min =     1
                                avg =    12.7
                                max =    20

                                Wald chi2(28)   =  2223.93
Log likelihood = -1376.386       Prob > chi2     =   0.0000

```

```

-----
                                spSE |      Coef.  Std. Err.   z    P>|z|    [95% Conf. Interval]
-----+-----
                                fallSE |  .5737737  .0167022   34.35  0.000    .5410379    .6065095
                                |
                                race |
                                Asian   |  .1371262  .1021901    1.34   0.180   -.0631626    .337415
                                African American | -.0241217  .075473   -0.32   0.749   -.1720462    .1238027
                                Multiracial | -.0014382  .070053   -0.02   0.984   -.1387397    .1358632
Native Hawaiian or Pacific Islander |  .2368842  .1417435    1.67   0.095   -.040928    .5146963
                                Unknown | -.1077341  .1221279   -0.88   0.378   -.3471003    .1316322
                                White  |  .0281118  .0554692    0.51   0.612   -.0806059    .1368294
                                |
                                language |
                                Amharic | .1990845  .2585316    0.77   0.441   -.3076281    .7057972
                                Arabic  | -.2601618  .1833602   -1.42   0.156   -.6195411    .0992175
                                Chinese | -.0390956  .27959    -0.14   0.889   -.587082    .5088907
                                English | -.0278451  .0952686   -0.29   0.770   -.2145682    .158878
                                Korean  |  .1870255  .4570156    0.41   0.682   -.7087086    1.08276

```

Kurdish		-.2151667	.6262837	-0.34	0.731	-1.44266	1.012327
Other		.0151612	.1520719	0.10	0.921	-.2828942	.3132167
Portuguese		-.0823463	.4461648	-0.18	0.854	-.9568131	.7921206
Punjabi		-.0982068	.2427484	-0.40	0.686	-.5739849	.3775713
Russian		-.2605543	.1698317	-1.53	0.125	-.5934183	.0723098
Spanish		.0608043	.1030609	0.59	0.555	-.1411915	.2628
Ukrainian		.0204019	.2098704	0.10	0.923	-.3909365	.4317403
Unknown		.3249095	.2670241	1.22	0.224	-.198448	.8482671
Vietnamese		-.0995069	.1541959	-0.65	0.519	-.4017252	.2027114
gender							
Male		-.1398933	.0337284	-4.15	0.000	-.2059996	-.0737869
Springagegrade							
Pre-K 4 class/grade		1.498104	.2192154	6.83	0.000	1.068449	1.927758
Preschool 3 class/grade		.9686866	.2181664	4.44	0.000	.5410883	1.396285
IEP							
Y		-.2351991	.0547739	-4.29	0.000	-.342554	-.1278441
classOrganization		.061192	.092296	0.66	0.507	-.1197049	.2420889
classEmotional		-.157372	.0900325	-1.75	0.080	-.3338324	.0190883
classInstruction		.0387029	.0599197	0.65	0.518	-.0787376	.1561434
_cons		2.840608	.4741606	5.99	0.000	1.91127	3.769946

```
-----
```

Random-effects Parameters		Estimate	Std. Err.	[95% Conf. Interval]
-----+				
ClassID: Identity				
var(_cons)		.2116578	.0341145	.1543262 .2902878
-----+				
var(Residual)		.3596674	.0141662	.3329469 .3885324

```
-----
```

LR test vs. linear regression: chibar2(01) = 394.06 Prob >= chibar2 = 0.0000

C.2 Physical regression

```

Mixed-effects ML regression           Number of obs   =    1400
Group variable: ClassID              Number of groups =    110

Obs per group: min =         1
                    avg =        12.7
                    max =         20

Wald chi2(28)      =    2021.76
Prob > chi2       =    0.0000

Log likelihood = -1078.0183

```

```

-----
                spPHYSICAL |      Coef.   Std. Err.   z   P>|z|   [95% Conf. Interval]
-----+-----
                fallPHYSICAL |   .5107269   .0176535   28.93   0.000   .4761267   .5453271
                |
                race |
                Asian |   .0189639   .0824245    0.23   0.818   -.1425851   .1805129
                African American |   .0149091   .0608112    0.25   0.806   -.1042787   .1340969
                Multiracial |  -.0232578   .0565013   -0.41   0.681   -.1339983   .0874826
                Native Hawaiian or Pacific Islander |   .0831616    .11439    0.73   0.467   -.1410386   .3073618
                Unknown |  -.1189113   .0985215   -1.21   0.227   -.31201    .0741873
                White |   .0058877   .0447323    0.13   0.895   -.0817861   .0935614
                |
                language |
                Amharic |   .1776155   .2083856    0.85   0.394   -.2308128   .5860439
                Arabic |  -.2152958   .147951   -1.46   0.146   -.5052744   .0746828
                Chinese |   .4684976   .2255557    2.08   0.038   .0264166   .9105786
                English |  -.0203402   .076595   -0.27   0.791   -.1704635   .1297832
                Korean |  -.2550813   .3688187   -0.69   0.489   -.9779527   .4677901
                Kurdish |   .6161545   .5037359    1.22   0.221   -.3711497   1.603459
                Other |   .1330369   .1225844    1.09   0.278   -.1072241   .3732979
                Portuguese |  -.087365   .3603139   -0.24   0.808   -.7935673   .6188373
                Punjabi |  -.4004643   .1951629   -2.05   0.040   -.7829766   -.017952
                Russian |  -.1166622   .1366215   -0.85   0.393   -.3844355   .1511111

```

Spanish		.0225441	.0828654	0.27	0.786	-.1398691	.1849573
Ukrainian		-.0754536	.1691769	-0.45	0.656	-.4070342	.2561271
Unknown		.4036508	.2157062	1.87	0.061	-.0191255	.8264271
Vietnamese		.0002746	.1243944	0.00	0.998	-.243534	.2440831
gender							
Male		-.032084	.0270622	-1.19	0.236	-.0851248	.0209569
Springagegrade							
Pre-K 4 class/grade		1.148528	.1639792	7.00	0.000	.8271351	1.469922
Preschool 3 class/grade		.5701451	.1626855	3.50	0.000	.2512873	.8890029
IEP							
Y		-.1823174	.0443023	-4.12	0.000	-.2691484	-.0954864
classOrganization		.1273024	.0725343	1.76	0.079	-.0148622	.2694669
classEmotional		-.166313	.0707101	-2.35	0.019	-.3049023	-.0277237
classInstruction		-.0202715	.0471189	-0.43	0.667	-.1126229	.0720799
_cons		3.606344	.373474	9.66	0.000	2.874348	4.338339

```
-----
-----
Random-effects Parameters | Estimate Std. Err. [95% Conf. Interval]
-----+-----
ClassID: Identity |
var(_cons) | .1294272 .0209005 .0943125 .1776159
-----+-----
var(Residual) | .2345592 .0092189 .2171689 .2533422
-----
```

LR test vs. linear regression: chibar2(01) = 372.00 Prob >= chibar2 = 0.0000

C.3 Language Regression

```
Mixed-effects ML regression          Number of obs   =   1386
Group variable: ClassID              Number of groups =   110

Obs per group: min =           1
                  avg =          12.6
```

max = 19

Wald chi2(28) = 3327.73
 Prob > chi2 = 0.0000
 Log likelihood = -1255.787

spLANGUAGE	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
fallLANGUAGE	.6298063	.0151643	41.53	0.000	.6000848	.6595279
race						
Asian	-.0429627	.0943085	-0.46	0.649	-.227804	.1418787
African American	-.0197694	.0695315	-0.28	0.776	-.1560487	.1165098
Multiracial	-.0267979	.0649367	-0.41	0.680	-.1540716	.1004757
Native Hawaiian or Pacific Islander	.1479527	.1284661	1.15	0.249	-.1038362	.3997416
Unknown	.0342237	.1126957	0.30	0.761	-.1866558	.2551032
White	.0074956	.0512316	0.15	0.884	-.0929166	.1079078
language						
Amharic	.056849	.2385134	0.24	0.812	-.4106286	.5243266
Arabic	-.3346387	.1687403	-1.98	0.047	-.6653636	-.0039138
Chinese	.115535	.257314	0.45	0.653	-.3887912	.6198612
English	-.005193	.0878838	-0.06	0.953	-.1774422	.1670562
Korean	.0602223	.4209582	0.14	0.886	-.7648406	.8852852
Kurdish	1.05539	.5779365	1.83	0.068	-.077345	2.188125
Other	.0282079	.1413763	0.20	0.842	-.2488845	.3053003
Portuguese	-.2653987	.4111211	-0.65	0.519	-1.071181	.5403839
Punjabi	.136552	.2229047	0.61	0.540	-.3003332	.5734372
Russian	-.2822828	.159804	-1.77	0.077	-.5954929	.0309273
Spanish	-.0689545	.0948191	-0.73	0.467	-.2547965	.1168874
Ukrainian	-.3388838	.1980696	-1.71	0.087	-.7270931	.0493256
Unknown	.5394251	.2461704	2.19	0.028	.05694	1.02191
Vietnamese	.0321851	.1428066	0.23	0.822	-.2477107	.3120809

gender							
Male		-.0680622	.0310481	-2.19	0.028	-.1289153	-.0072091
Springagegrade							
Pre-K 4 class/grade		1.32417	.1868042	7.09	0.000	.9580409	1.6903
Preschool 3 class/grade		.8736661	.1862277	4.69	0.000	.5086664	1.238666
IEP							
Y		-.3723033	.0516465	-7.21	0.000	-.4735286	-.2710779
classOrganization		.1046987	.0875934	1.20	0.232	-.0669813	.2763787
classEmotional		-.2095559	.0853099	-2.46	0.014	-.3767603	-.0423515
classInstruction		.0165059	.0567589	0.29	0.771	-.0947395	.1277513
_cons		2.834701	.4399453	6.44	0.000	1.972425	3.696978

```

-----
-----
Random-effects Parameters | Estimate Std. Err. [95% Conf. Interval]
-----+-----
ClassID: Identity |
var(_cons) | .1917378 .0310255 .1396282 .2632949
-----+-----
var(Residual) | .3051214 .0120748 .2823498 .3297295
-----

```

LR test vs. linear regression: chibar2(01) = 400.07 Prob >= chibar2 = 0.0000

C.4 Literacy Regression

Mixed-effects ML regression

```

Number of obs = 1323
Group variable: ClassID
Number of groups = 107

Obs per group: min = 1
                avg = 12.4
                max = 20

Wald chi2(28) = 3121.70
Log likelihood = -1371.2229
Prob > chi2 = 0.0000

```

spLITERACY	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
fallLITERACY	.7920938	.0197939	40.02	0.000	.7532985	.8308891
race						
Asian	-.0351628	.1089854	-0.32	0.747	-.2487703	.1784447
African American	.1074895	.0808601	1.33	0.184	-.0509935	.2659724
Multiracial	-.0191653	.0753694	-0.25	0.799	-.1668866	.128556
Native Hawaiian or Pacific Islander	.1985368	.1604673	1.24	0.216	-.1159733	.5130469
Unknown	.0638115	.1301639	0.49	0.624	-.191305	.318928
White	.0515017	.0602119	0.86	0.392	-.0665114	.1695148
language						
Amharic	.316086	.3386834	0.93	0.351	-.3477213	.9798934
Arabic	-.2608569	.1915082	-1.36	0.173	-.636206	.1144922
Chinese	.1432756	.2936157	0.49	0.626	-.4322005	.7187517
English	-.0207978	.1004513	-0.21	0.836	-.2176787	.1760832
Korean	.2274618	.4771614	0.48	0.634	-.7077574	1.162681
Kurdish	1.037593	.6569772	1.58	0.114	-.250059	2.325244
Other	.2818354	.1611062	1.75	0.080	-.033927	.5975978
Portuguese	-.4880678	.4653345	-1.05	0.294	-1.400107	.423971
Punjabi	.1945091	.2556075	0.76	0.447	-.3064723	.6954906
Russian	-.2575041	.1839644	-1.40	0.162	-.6180676	.1030594
Spanish	-.0849084	.1090836	-0.78	0.436	-.2987084	.1288916
Ukrainian	-.5212881	.2348106	-2.22	0.026	-.9815085	-.0610677
Unknown	.1718593	.2787157	0.62	0.537	-.3744134	.718132
Vietnamese	.0308476	.1629667	0.19	0.850	-.2885613	.3502565
gender						
Male	.0169989	.0362179	0.47	0.639	-.0539869	.0879848
Springagegrade						
Pre-K 4 class/grade	1.499498	.2580456	5.81	0.000	.9937379	2.005258

Preschool 3 class/grade		.9328513	.2553149	3.65	0.000	.4324433	1.433259
IEP							
Y		-.3623683	.0592418	-6.12	0.000	-.4784802	-.2462564
classOrganization		.1314422	.108253	1.21	0.225	-.0807297	.3436141
classEmotional		-.2215688	.105895	-2.09	0.036	-.4291192	-.0140184
classInstruction		-.0425948	.0709914	-0.60	0.549	-.1817354	.0965457
_cons		1.796681	.5515322	3.26	0.001	.7156972	2.877664

```

-----
Random-effects Parameters | Estimate   Std. Err.   [95% Conf. Interval]
-----+-----
ClassID: Identity       |
    var(_cons)          |   .2962883   .048363    .2151659   .4079959
-----+-----
    var(Residual)       |   .3903309   .0158493   .3604708   .4226645
-----

```

LR test vs. linear regression: chibar2(01) = 415.18 Prob >= chibar2 = 0.0000

C.5 Cognitive Regression

Mixed-effects ML regression

Number of obs	=	1390
Group variable: ClassID		
Number of groups	=	110
Obs per group: min	=	1
avg	=	12.6
max	=	20

Wald chi2(28)	=	2884.66
Log likelihood = -1171.5162		
Prob > chi2	=	0.0000

```

-----
spCOGNITIVE |      Coef.   Std. Err.      z    P>|z|    [95% Conf. Interval]
-----+-----
fallCOGNITIVE |   .6586907   .0174524   37.74   0.000   .6244846   .6928968
-----

```

	race						
Asian		-.0776697	.0878587	-0.88	0.377	-.2498697	.0945302
African American		-.0140592	.0645242	-0.22	0.828	-.1405243	.1124058
Multiracial		-.068299	.060025	-1.14	0.255	-.1859458	.0493479
Native Hawaiian or Pacific Islander		.1079811	.1211488	0.89	0.373	-.1294662	.3454284
Unknown		-.0531229	.1044844	-0.51	0.611	-.2579086	.1516627
White		.0037169	.0473843	0.08	0.937	-.0891546	.0965883
	language						
Amharic		-.141042	.2234909	-0.63	0.528	-.579076	.296992
Arabic		-.3672046	.1568959	-2.34	0.019	-.6747149	-.0596944
Chinese		.3257617	.2393205	1.36	0.173	-.1432979	.7948213
English		-.0189823	.0817769	-0.23	0.816	-.1792621	.1412975
Korean		.1078767	.3906863	0.28	0.782	-.6578544	.8736077
Kurdish		.976302	.5344823	1.83	0.068	-.0712642	2.023868
Other		.1622323	.1314919	1.23	0.217	-.0954871	.4199517
Portuguese		-.098897	.3810923	-0.26	0.795	-.8458242	.6480302
Punjabi		.3289699	.2074991	1.59	0.113	-.0777208	.7356607
Russian		-.2373528	.1450407	-1.64	0.102	-.5216274	.0469217
Spanish		-.0376447	.0883341	-0.43	0.670	-.2107764	.1354869
Ukrainian		-.16133	.1793611	-0.90	0.368	-.5128714	.1902113
Unknown		.2970515	.2283585	1.30	0.193	-.1505229	.7446259
Vietnamese		.10964	.1323072	0.83	0.407	-.1496773	.3689573
	gender						
Male		-.0724833	.0289016	-2.51	0.012	-.1291293	-.0158373
	Springagegrade						
Pre-K 4 class/grade		1.527327	.1800143	8.48	0.000	1.174505	1.880148
Preschool 3 class/grade		1.021297	.1787115	5.71	0.000	.6710289	1.371565
	IEP						
Y		-.2352494	.0473977	-4.96	0.000	-.3281473	-.1423515
classOrganization		.0875139	.0964848	0.91	0.364	-.1015927	.2766206

classEmotional		-.215645	.0941156	-2.29	0.022	-.4001082	-.0311819
classInstruction		.0153281	.0622715	0.25	0.806	-.1067218	.137378
_cons		2.50754	.46867	5.35	0.000	1.588964	3.426116

```
-----
Random-effects Parameters | Estimate Std. Err. [95% Conf. Interval]
-----+-----
ClassID: Identity |
var(_cons) | .2439229 .037354 .1806761 .3293096
-----+-----
var(Residual) | .2618928 .010344 .2423839 .282972
-----
```

LR test vs. linear regression: $\chi^2(01) = 616.95$ Prob $\geq \chi^2 = 0.0000$

C.6 Mathematics regression

Mixed-effects ML regression

Number of obs	=	1186
Number of groups	=	103
Obs per group: min	=	1
avg	=	11.5
max	=	19

Wald $\chi^2(27)$	=	2434.35
Log likelihood = -1288.8079	Prob > χ^2	= 0.0000

```
-----
spMATH | Coef. Std. Err. z P>|z| [95% Conf. Interval]
-----+-----
fallMATH | .6836782 .0199385 34.29 0.000 .6445994 .722757
|
race |
Asian | .0154823 .1197645 0.13 0.897 -.2192518 .2502164
African American | .090756 .0918618 0.99 0.323 -.0892897 .2708017
Multiracial | -.0310687 .0839343 -0.37 0.711 -.1955768 .1334394
-----
```

Appendix

Native Hawaiian or Pacific Islander		.359771	.1771385	2.03	0.042	.012586	.7069561
Unknown		-.1470283	.1468937	-1.00	0.317	-.4349347	.1408781
White		.0125687	.0675409	0.19	0.852	-.1198091	.1449465
language							
Amharic		.2040025	.3594803	0.57	0.570	-.5005659	.9085709
Arabic		.1499984	.2167026	0.69	0.489	-.2747308	.5747276
Chinese		.1994264	.3367555	0.59	0.554	-.4606023	.8594552
English		-.0255657	.1111627	-0.23	0.818	-.2434406	.1923091
Korean		.0447149	.5055622	0.09	0.930	-.9461689	1.035599
Other		.2978966	.1755684	1.70	0.090	-.0462111	.6420043
Portuguese		-.0910825	.4930207	-0.18	0.853	-1.057385	.8752204
Punjabi		-.1595119	.2696803	-0.59	0.554	-.6880756	.3690518
Russian		-.094622	.2056674	-0.46	0.645	-.4977227	.3084786
Spanish		-.013164	.1212946	-0.11	0.914	-.250897	.2245689
Ukrainian		-.5104277	.2343313	-2.18	0.029	-.9697086	-.0511468
Unknown		.1476597	.3538974	0.42	0.677	-.5459664	.8412858
Vietnamese		.1075025	.1794737	0.60	0.549	-.2442594	.4592644
gender							
Male		.0310735	.0401255	0.77	0.439	-.047571	.1097181
Springagegrade							
Pre-K 4 class/grade		1.879317	.271818	6.91	0.000	1.346564	2.412071
Preschool 3 class/grade		1.203756	.2686977	4.48	0.000	.6771177	1.730393
IEP							
Y		-.3717171	.0671268	-5.54	0.000	-.5032833	-.240151
classOrganization		.1151577	.1080604	1.07	0.287	-.0966369	.3269523
classEmotional		-.162472	.1042114	-1.56	0.119	-.3667225	.0417785
classInstruction		-.0093049	.0691422	-0.13	0.893	-.1448212	.1262113
_cons		1.572388	.5623766	2.80	0.005	.4701503	2.674626

```

Random-effects Parameters | Estimate   Std. Err.   [95% Conf. Interval]
-----+-----
ClassID: Identity        |
      var(_cons) |   .2679363   .0455855   .1919604   .3739825
-----+-----
      var(Residual) |   .436307   .0187109   .4011333   .4745649
-----+-----

LR test vs. linear regression: chibar2(01) =   351.01 Prob >= chibar2 = 0.0000

```

Appendix D Ordinal logistical regressions

D.1 Social Emotional

```

Ordered logistic regression           Number of obs   =       1396
                                      LR chi2(28)      =       728.48
                                      Prob > chi2     =       0.0000
Log likelihood = -900.88872           Pseudo R2      =       0.2879

```

```

-----+-----
      springSECat |      Coef.   Std. Err.      z    P>|z|    [95% Conf. Interval]
-----+-----
      fallSECat |   .1039375   .0062849    16.54  0.000    .0916193    .1162557
      |
      race |
      Asian |   .4543826   .4142231     1.10  0.273   -.3574797    1.266245
      African American |   .1888716   .2939826     0.64  0.521   -.3873237    .7650669
      Multiracial |   .0208968   .2730058     0.08  0.939   -.5141847    .5559784
      Native Hawaiian or Pacific Islander |   1.323241   .599249     2.21  0.027    .1487341    2.497747
      Unknown |   -.377803   .5460282    -0.69  0.489   -1.447999    .6923926
      White |   .1006066   .2233361     0.45  0.652   -.3371241    .5383373
      |
      language |
      Amharic |  -.0404042   .931047    -0.04  0.965   -1.865223    1.784414
      Arabic |   .2669445   .8206573     0.33  0.745   -1.341514    1.875403
      Chinese |   .0362796   1.367082     0.03  0.979   -2.643151    2.715711
      English |  -.1165469   .340055    -0.34  0.732   -.7830425    .5499487
      Korean |   .3903634   1.411054     0.28  0.782   -2.375252    3.155979

```

Kurdish		1.947895	1.711994	1.14	0.255	-1.407551	5.303341
Other		-.3254845	.5904287	-0.55	0.581	-1.482703	.8317345
Portuguese		11.7108	525.5548	0.02	0.982	-1018.358	1041.779
Punjabi		-.4085513	.9493931	-0.43	0.667	-2.269328	1.452225
Russian		-.1787791	.6182086	-0.29	0.772	-1.390446	1.032888
Spanish		.5381878	.378246	1.42	0.155	-.2031607	1.279536
Ukrainian		.2234116	.8837569	0.25	0.800	-1.50872	1.955543
Unknown		.5465066	1.200156	0.46	0.649	-1.805756	2.898769
Vietnamese		-.0872627	.5867846	-0.15	0.882	-1.237339	1.062814
gender							
Male		-.5503284	.1347787	-4.08	0.000	-.8144898	-.2861669
Springagegrade							
Pre-K 4 class/grade		3.325412	.608239	5.47	0.000	2.133286	4.517538
Preschool 3 class/grade		1.537377	.6042606	2.54	0.011	.3530475	2.721706
IEP							
Y		-.2335154	.2041125	-1.14	0.253	-.6335685	.1665377
classOrganization		-.0842496	.126412	-0.67	0.505	-.3320126	.1635134
classEmotional		-.4305625	.1290409	-3.34	0.001	-.683478	-.1776469
classInstruction		.2548497	.0842525	3.02	0.002	.0897179	.4199815

/cut1		-3.003904	.9998113			-4.963498	-1.04431
/cut2		-1.112347	.9293509			-2.933842	.7091471
/cut3		1.567115	.9229059			-.2417469	3.375978
/cut4		3.884887	.9304443			2.061249	5.708524

D.2 Physical**Ordered logistic regression**

Number of obs	=	1389
LR chi2(28)	=	672.82
Prob > chi2	=	0.0000
Pseudo R2	=	0.2948
Log likelihood = -804.73522		

springPhysicalCat	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
-------------------	-------	-----------	---	------	----------------------

fallPhysicalCat		.1721552	.0125271	13.74	0.000	.1476026	.1967078
race							
Asian		.6798801	.4378442	1.55	0.120	-.1782788	1.538039
African American		.2003751	.3027633	0.66	0.508	-.3930299	.7937802
Multiracial		.2967832	.2837863	1.05	0.296	-.2594277	.8529942
Native Hawaiian or Pacific Islander		-.1212091	.5384466	-0.23	0.822	-1.176545	.9341269
Unknown		-.6722872	.4822412	-1.39	0.163	-1.617463	.2728882
White		.0843897	.2270875	0.37	0.710	-.3606936	.5294729
language							
Amharic		.9694593	.9488885	1.02	0.307	-.890328	2.829247
Arabic		.3223866	.7568024	0.43	0.670	-1.160919	1.805692
Chinese		16.21111	713.5189	0.02	0.982	-1382.26	1414.682
English		.3746487	.3484666	1.08	0.282	-.3083333	1.057631
Korean		.5783927	1.595715	0.36	0.717	-2.549151	3.705937
Kurdish		15.05135	1711.044	0.01	0.993	-3338.533	3368.636
Other		.8118615	.637597	1.27	0.203	-.4378057	2.061529
Portuguese		12.8575	1191.273	0.01	0.991	-2321.995	2347.71
Punjabi		-1.512701	.8182425	-1.85	0.064	-3.116427	.0910248
Russian		-.1991174	.6474036	-0.31	0.758	-1.468005	1.06977
Spanish		.8308023	.3871471	2.15	0.032	.0720079	1.589597
Ukrainian		.0723172	.8857781	0.08	0.935	-1.663776	1.80841
Unknown		14.92631	617.8855	0.02	0.981	-1196.107	1225.96
Vietnamese		-.3735428	.5902334	-0.63	0.527	-1.530379	.7832935
gender							
Male		-.2639478	.1367419	-1.93	0.054	-.531957	.0040615
Springagegrade							
Pre-K 4 class/grade		3.328514	.5582818	5.96	0.000	2.234302	4.422726
Preschool 3 class/grade		1.192742	.5530267	2.16	0.031	.1088301	2.276655
IEP							

Y		-.5310125	.2037195	-2.61	0.009	-.9302953	-.1317296
classOrganization		.4882729	.1346727	3.63	0.000	.2243192	.7522265
classEmotional		-.4769099	.1394998	-3.42	0.001	-.7503245	-.2034953
classInstruction		-.1002663	.0874849	-1.15	0.252	-.2717336	.0712011

/cut1		-2.394303	1.350389			-5.041017	.2524115
/cut2		.5256377	.9514186			-1.339108	2.390384
/cut3		3.090224	.9352482			1.257171	4.923277
/cut4		6.239248	.9591135			4.35942	8.119076

D.3 Language

Ordered logistic regression

Number of obs = 1373

LR chi2(28) = 937.94

Prob > chi2 = 0.0000

Log likelihood = -990.90205

Pseudo R2 = 0.3212

springLanguageCat		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
fallLanguageCat		.1241716	.0065282	19.02	0.000	.1113766 .1369667
race						
Asian		.0983915	.3722271	0.26	0.792	-.6311603 .8279433
African American		.2138305	.2748118	0.78	0.437	-.3247907 .7524518
Multiracial		-.0811867	.2534851	-0.32	0.749	-.5780083 .415635
Native Hawaiian or Pacific Islander		.1433605	.5045205	0.28	0.776	-.8454816 1.132203
Unknown		.3606608	.4467164	0.81	0.419	-.5148873 1.236209
White		.0018449	.1954129	0.01	0.992	-.3811573 .3848472
language						
Amharic		-1.18621	.8450978	-1.40	0.160	-2.842571 .4701514
Arabic		-.3968827	.6888447	-0.58	0.565	-1.746994 .9532281
Chinese		-.2784764	1.004312	-0.28	0.782	-2.246892 1.689939
English		.0553983	.3360711	0.16	0.869	-.6032889 .7140854
Korean		.5821752	1.35487	0.43	0.667	-2.073322 3.237672

Kurdish		-.5771851	1.745605	-0.33	0.741	-3.998507	2.844137
Other		.3707461	.5689215	0.65	0.515	-.7443196	1.485812
Portuguese		-1.355806	1.58432	-0.86	0.392	-4.461017	1.749404
Punjabi		-.416896	.8575372	-0.49	0.627	-2.097638	1.263846
Russian		-1.576912	.5812755	-2.71	0.007	-2.716191	-.4376325
Spanish		-.1937514	.3646442	-0.53	0.595	-.9084408	.5209381
Ukrainian		-1.29131	.6840712	-1.89	0.059	-2.632065	.0494453
Unknown		15.18266	795.9932	0.02	0.985	-1544.935	1575.301
Vietnamese		-.1130863	.5506903	-0.21	0.837	-1.192419	.9662468
gender							
Male		-.3195445	.1230027	-2.60	0.009	-.5606255	-.0784636
Springagegrade							
Pre-K 4 class/grade		3.989574	.5427469	7.35	0.000	2.92581	5.053338
Preschool 3 class/grade		2.318449	.5388906	4.30	0.000	1.262243	3.374655
IEP							
Y		-.9062324	.1899586	-4.77	0.000	-1.278544	-.5339204
classOrganization		.1938274	.1191351	1.63	0.104	-.0396732	.4273279
classEmotional		-.4587377	.1172112	-3.91	0.000	-.6884674	-.2290079
classInstruction		.0321086	.0775711	0.41	0.679	-.119928	.1841451

/cut1		-1.814948	.8999979			-3.578912	-.0509846
/cut2		.4504573	.8490534			-1.213657	2.114571
/cut3		2.948778	.8566678			1.26974	4.627816
/cut4		6.157201	.8772466			4.437829	7.876572

D.4 Literacy

Ordered logistic regression

Number of obs	=	1323
LR chi2(28)	=	850.51
Prob > chi2	=	0.0000
Pseudo R2	=	0.3395

Log likelihood = -827.3974

springLiteracyCat	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
fallLiteracyCat	.1435192	.0082687	17.36	0.000	.127313	.1597255
race						
Asian	.8354898	.4358598	1.92	0.055	-.0187796	1.689759
African American	.3586139	.295383	1.21	0.225	-.2203261	.937554
Multiracial	-.0433169	.2697437	-0.16	0.872	-.5720048	.4853709
Native Hawaiian or Pacific Islander	.5514852	.5759264	0.96	0.338	-.5773099	1.68028
Unknown	.1087701	.489061	0.22	0.824	-.8497718	1.067312
White	.0020692	.2128374	0.01	0.992	-.4150845	.4192229
language						
Amharic	.5943017	1.208648	0.49	0.623	-1.774605	2.963208
Arabic	.2726051	.7550193	0.36	0.718	-1.207206	1.752416
Chinese	10.6008	393.8399	0.03	0.979	-761.3112	782.5128
English	-.2039008	.3711547	-0.55	0.583	-.9313506	.523549
Korean	.005901	1.46454	0.00	0.997	-2.864545	2.876347
Kurdish	.9898603	2.113449	0.47	0.640	-3.152424	5.132145
Other	-.4968069	.601317	-0.83	0.409	-1.675367	.6817526
Portuguese	-2.02287	1.874979	-1.08	0.281	-5.697761	1.652022
Punjabi	-.6137746	.8688187	-0.71	0.480	-2.316628	1.089079
Russian	-.7436829	.6448089	-1.15	0.249	-2.007485	.5201193
Spanish	.0151018	.4024845	0.04	0.970	-.7737532	.8039569
Ukrainian	-1.228217	.7906162	-1.55	0.120	-2.777796	.3213624
Unknown	-.4508439	1.182506	-0.38	0.703	-2.768513	1.866825
Vietnamese	-1.004615	.6188193	-1.62	0.104	-2.217478	.2082489
gender						
Male	-.0295916	.1343253	-0.22	0.826	-.2928643	.2336811
Springagegrade						
Pre-K 4 class/grade	3.519074	.6626168	5.31	0.000	2.220369	4.817779
Preschool 3 class/grade	1.831584	.6524504	2.81	0.005	.5528049	3.110363

IEP							
Y		-.7831467	.203107	-3.86	0.000	-1.181229	-.3850643
classOrganization		.0921457	.1288742	0.72	0.475	-.1604432	.3447346
classEmotional		-.4012724	.1295532	-3.10	0.002	-.655192	-.1473527
classInstruction		-.0712225	.0889407	-0.80	0.423	-.245543	.1030981

/cut1		-3.994822	1.091556			-6.134233	-1.855412
/cut2		-1.289133	.9636463			-3.177845	.5995785
/cut3		.5186376	.9617937			-1.366443	2.403719
/cut4		4.226315	.975581			2.314211	6.138418

D.5 Cognitive

Ordered logistic regression

Number of obs	=	1344
LR chi2(28)	=	706.08
Prob > chi2	=	0.0000
Pseudo R2	=	0.3181

Log likelihood = -756.6359

springCognitiveCat		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
fallCognitiveCat		.133398	.0085461	15.61	0.000	.1166481 .150148
race						
Asian		.4919664	.4361083	1.13	0.259	-.3627901 1.346723
African American		.4333941	.3008269	1.44	0.150	-.1562157 1.023004
Multiracial		.159839	.2760669	0.58	0.563	-.3812422 .7009203
Native Hawaiian or Pacific Islander		1.783523	.6776812	2.63	0.008	.4552919 3.111753
Unknown		.006737	.5187757	0.01	0.990	-1.010045 1.023519
White		.2205586	.2182688	1.01	0.312	-.2072404 .6483575
language						
Amharic		-.3206799	.9729264	-0.33	0.742	-2.227581 1.586221
Arabic		.5508388	.8577323	0.64	0.521	-1.130286 2.231963
Chinese		13.7709	940.5329	0.01	0.988	-1829.64 1857.182
English		-.1237346	.3673704	-0.34	0.736	-.8437673 .5962981

Korean		1.451085	1.908734	0.76	0.447	-2.289965	5.192135
Kurdish		2.159872	2.241294	0.96	0.335	-2.232985	6.552728
Other		.3227787	.6403409	0.50	0.614	-.9322664	1.577824
Portuguese		14.1625	1625.639	0.01	0.993	-3172.031	3200.356
Punjabi		.3581145	1.096873	0.33	0.744	-1.791718	2.507947
Russian		-1.016073	.636785	-1.60	0.111	-2.264149	.2320027
Spanish		.1583532	.4023237	0.39	0.694	-.6301867	.9468931
Ukrainian		-.5484454	.817962	-0.67	0.503	-2.151621	1.054731
Unknown		1.00491	1.368414	0.73	0.463	-1.677133	3.686953
Vietnamese		.3181191	.6314028	0.50	0.614	-.9194076	1.555646
gender							
Male		-.2590521	.1367061	-1.89	0.058	-.5269911	.0088869
Springagegrade							
Pre-K 4 class/grade		4.332377	.6453764	6.71	0.000	3.067462	5.597292
Preschool 3 class/grade		2.545418	.6377845	3.99	0.000	1.295383	3.795453
IEP							
Y		-.60233	.2064514	-2.92	0.004	-1.006967	-.1976928
classOrganization		.1596509	.1317345	1.21	0.226	-.098544	.4178457
classEmotional		-.4892235	.1331322	-3.67	0.000	-.7501577	-.2282892
classInstruction		.139975	.0858054	1.63	0.103	-.0282005	.3081505

/cut1		.0078888	1.010949			-1.973535	1.989313
/cut2		3.084832	.980629			1.162835	5.00683
/cut3		7.094045	1.016254			5.102224	9.085865

D.6 Mathematics

Ordered logistic regression

Number of obs	=	1183
LR chi2(27)	=	811.69
Prob > chi2	=	0.0000
Log likelihood = -969.9194	Pseudo R2	= 0.2950

Appendix

springMathCat	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
fallMathCat	.1781153	.0096719	18.42	0.000	.1591588	.1970719
race						
Asian	.4707698	.3742945	1.26	0.208	-.262834	1.204374
African American	.3634106	.2759063	1.32	0.188	-.1773558	.904177
Multiracial	-.0069825	.2543285	-0.03	0.978	-.5054573	.4914922
Native Hawaiian or Pacific Islander	1.05113	.5620281	1.87	0.061	-.0504247	2.152685
Unknown	-.5078216	.4400522	-1.15	0.248	-1.370308	.3546648
White	-.0308298	.2022111	-0.15	0.879	-.4271562	.3654965
language						
Amharic	-.1045026	.9954317	-0.10	0.916	-2.055513	1.846508
Arabic	.5304005	.7499848	0.71	0.479	-.9395428	2.000344
Chinese	-.2167366	1.240237	-0.17	0.861	-2.647557	2.214083
English	.0092689	.3253109	0.03	0.977	-.6283288	.6468666
Korean	-1.054465	1.40612	-0.75	0.453	-3.810409	1.701479
Other	.2461327	.5425669	0.45	0.650	-.8172788	1.309544
Portuguese	1.100555	1.702705	0.65	0.518	-2.236685	4.437796
Punjabi	-1.505983	.8191947	-1.84	0.066	-3.111576	.0996087
Russian	-.5161974	.6172137	-0.84	0.403	-1.725914	.6935191
Spanish	.0470063	.3577001	0.13	0.895	-.6540729	.7480856
Ukrainian	-1.618392	.6499809	-2.49	0.013	-2.892331	-.3444527
Unknown	-.6578391	1.02896	-0.64	0.523	-2.674563	1.358885
Vietnamese	.0569671	.5760206	0.10	0.921	-1.072013	1.185947
gender						
Male	-.1298584	.1242549	-1.05	0.296	-.3733936	.1136768
Springagegrade						
Pre-K 4 class/grade	3.767611	.6902176	5.46	0.000	2.414809	5.120412
Preschool 3 class/grade	2.163294	.6843326	3.16	0.002	.8220266	3.504561
IEP						

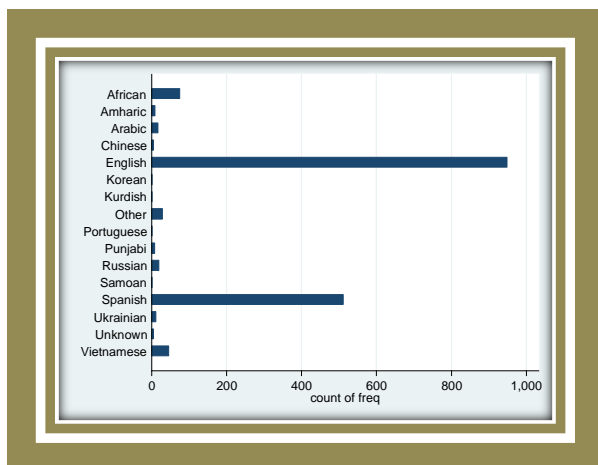
Y		-.4421977	.2038387	-2.17	0.030	-.8417143	-.0426811
classOrganization		.0578783	.1170498	0.49	0.621	-.1715352	.2872917
classEmotional		-.1996905	.1103123	-1.81	0.070	-.4158986	.0165177
classInstruction		.0403264	.0766471	0.53	0.599	-.1098992	.1905519

/cut1		-1.424731	1.00701			-3.398434	.5489723
/cut2		.9635801	.9380681			-.8749997	2.80216
/cut3		4.16547	.9516477			2.300275	6.030665
/cut4		7.33156	.9651098			5.439979	9.22314

Appendix E Population Descriptive

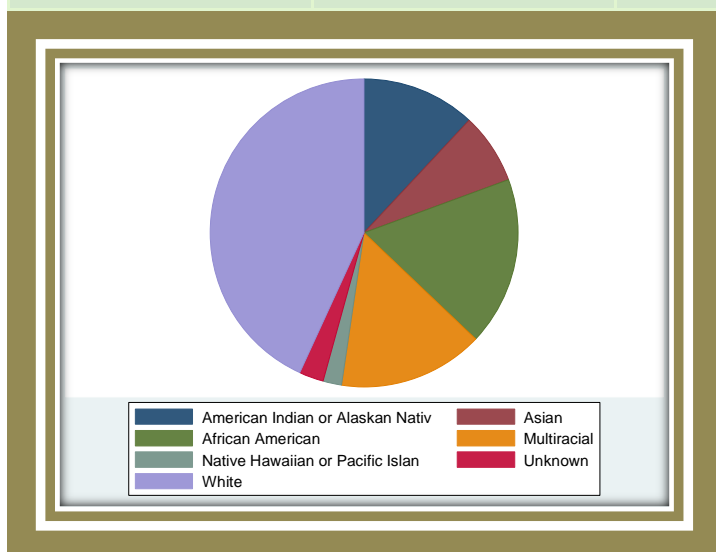
E.1 Language Table and Bar Graph

Language		Freq.	Percent	Cum.
African		75	4.42	4.42
Amharic		10	0.59	5.01
Arabic		18	1.06	6.07
Chinese		6	0.35	6.42
English		948	55.83	62.25
Korean		2	0.12	62.37
Kurdish		1	0.06	62.43
Other		30	1.77	64.19
Portuguese		2	0.12	64.31
Punjabi		9	0.53	64.84
Russian		20	1.18	66.02
Samoan		1	0.06	66.08
Spanish		512	30.15	96.23
Ukrainian		12	0.71	96.94
Unknown		6	0.35	97.29
Vietnamese		46	2.71	100.00
Total		1,698	100.00	



E.2 Race Table and Pie Chart

Race	Frequency	Percent
American Indian or Alaskan Native	203	11.96%
Asian	126	7.42%
African American	301	17.73%
Multiracial	259	15.52%
Native Hawaiian or Pacific Islander	33	1.93%
Unknown	44	2.59%
Caucasian	732	43.11%



E.3 Age Group Table

Spring Age Grade	Frequency	Percent
2-3	26	1.61%
Preschool 3	429	26.51%
Pre-K4	1,161	71.76%
Kindergarten	2	.12%
Total	1,618	