Observing Individual Children in Early Childhood Classrooms Using Optimizing Learning

Opportunities for Students (OLOS): A Feasibility Study

Carol McDonald Connor[1]

Ashley Adams[1]

Elham Zargar[1]

Taffeta S. Wood[1]

Belinda E. Hernandez[1]

Deborah Lowe Vandell[1]

Affiliations:
[1]University of California, Irvine

Acknowledgements

Abstract

In this feasibility study, we present a newly developed observational system, Optimizing Learning Opportunities for Students (OLOS). OLOS is designed to elucidate the learning opportunities afforded to individual children within early childhood classrooms and as they transition to formal schooling (kindergarten through third grade). OLOS records the time spent in different types of learning opportunities (e.g., play, literacy, math) and the frequency of specific discourse moves children and teachers use (child talk and teacher talk). Importantly, it is designed to be used validly and reliably by practitioners. Using OLOS, we explored individual children's experiences (n = 68 children in 12 classrooms) in four different types of early childhood programs; state-funded pre-kindergarten (PK), state-funded PK serving children with disabilities, Head Start, and a tuition-based (non-profit) preschool. Results of our feasibility study revealed that coders achieved adequate inter-rater reliability with about 10 hours of training and 5 hours of practice. We could feasibly and reliably use OLOS in these very different kinds of pre-kindergarten programs but with some changes. In analyzing the observations, we found that individual children's learning opportunities varied significantly both within and between classrooms. In general, we observed that most of the PK day (or half day) was spent in language and literacy activities and non-instructional activities (e.g., transitions). Very little time in math and science was observed yet children were generally more likely to actively participate (i.e., more child talk) during academic learning opportunities (literacy, math, and science). The frequency of teacher talk also varied widely between classrooms and across programs. Plus, the more teacher talk we observed, the more child talk we observed. Our long-term aim is that

OLOS can inform policy and provide information that supports practitioners in meeting the

learning and social-behavioral needs of the children they serve.

Observing Individual Children in Early Childhood Classrooms using

Optimizing Learning Opportunities for Students (OLOS): A Feasibility Study

An important advance in early childhood research and policy has been the development of observation tools to assess early childhood learning environments. In line with the varied needs within the field, several observation tools have been developed to serve a variety of purposes including program accreditation (Bredekamp, 1986), program improvement and staff development (Pianta & Hamre, 2009), evaluation of program impacts (Burchinal, 2018), and measures of child competencies within a classroom context (Downer, Booren, Lima, Luckner, & Pianta, 2010). However, there are few observation tools that have been designed specifically for practitioners to use. Also, current tools have primarily assessed overall classroom quality and few tools have measured the experiences of multiple individual children simultaneously within a classroom. The purpose of this paper is to present a feasibility study of a new, low-inference observation system, *Optimizing Learning Opportunities for Students* (OLOS) that is designed to observe multiple individual children's learning opportunities in the classroom from pre-kindergarten (PK) through third grade, focusing, for this study, specifically on PK classrooms. By low inference, we mean that OLOS relies on timing (in seconds) how much of what types of instruction are provided and counting the frequency of different types of teacher and student talk. There are no Likert-rating scales or rubrics that require extensive training for calibration.

Existing Measures of Classroom Quality

Among the assessments of overall early childhood classroom quality, the *Classroom Assessment Scoring System* (CLASS; Pianta & Hamre, 2009), the *Early Childhood Environment*

Rating System (ECERS; Harms, Clifford, & Cryer, 2015), and the *Early Childhood Classroom Observation Measure* (ECCOM; Stipek & Byler, 2004) are among the most well-known (Burchinal, 2018). These observational tools have been widely and successfully used in diverse early education settings including Head Start, state funded PK programs, and community-based early childhood centers. They provide complementary descriptions of different aspects of the overall process quality of early childhood classrooms while using different protocols. For the CLASS, trained observers focus on teachers and how they interact with children in their classrooms and use global ratings to evaluate three dimensions: instructional support, emotional support, and classroom management. For the ECERS, trained observers rate aspects of early childhood settings such as space and furnishings, personal care routines, and learning activities using 7-point scales. For the ECCOM, trained observers assess the degree to which instructional practices are constructivist, child-centered, and the degree to which didactic, teacher-centered instructional practices are implemented.

A primary motivation underlying these classroom ratings are efforts to ensure that classrooms are, in general, providing experiences that foster children's learning and development. As such, they focus on teachers' efforts to organize the classroom as a whole. The CLASS and the ECERS have been used to guide quality improvement efforts by identifying aspects of teacher-child interactions, the physical space, and teacher behavior that might be improved (Hestenes et al., 2015). Concerns have been expressed that global assessments, while informative, do not examine the experiences of the individual children in the classroom (Burchinal, 2018; Downer, Booren, Lima, Luckner, & Pianta, 2010). Activities may be well-matched to the skills and characteristics of some children in the classroom, but not well-aligned

with the skills of other children in the same classroom. OLOS, by coding the different learning opportunities afforded to multiple individual children was designed to fill this gap. By design, data collected using OLOS should help teachers better individualize or personalize the learning opportunities afforded to the children in their classrooms.

Other observation systems have focused on the learning opportunities, behaviors, and experiences of individual children in early childhood settings. For example, the *Observational Record of the Caregiving Environment* (ORCE; NICHD Early Child Care Research Network, 2000) uses time-sampled behavioral counts of teacher behaviors (e.g., positive talk, read aloud, respond to child's vocalizations) with individual study children as well as 4-point ratings of teachers' sensitivity-responsivity, warmth, and stimulation with individual study children. A second measure, the Emerging Academic Snapshot (Richie, Howes, Kraft-Sayre, & Weiser, 2001), is a time-sample coding system (20-seconds observe and 40-seconds record) that measures individual children's time in different activity settings (i.e., whole class, interactions with adults, engagement in academic activities, waiting in line, cleanup, and toileting), which are then aggregated across the classroom. A third measure, the *Individualized Classroom Assessment Scoring System* (inCLASS; Downer et al., 2010), rates individual children's interactions in preschool classrooms with their teachers and peers, and their engagement in classroom activities in order to assess child competencies in domains such as task orientation, peer interactions, and teacher-child interactions. OLOS is designed to observe multiple individual children simultaneously in order to describe and understand the variations in learning opportunities within a single classroom.

Optimizing Learning Opportunities for Students

Here, we introduce a new early childhood observational assessment, *Optimizing Learning Opportunities for Students* (OLOS). Designed to be low-inference so it can be used by practitioners, the primary goal of OLOS is to enable teachers and coaches to better understand and meet the learning needs of children from age four (PK) through third grade. We do this by carefully assessing individual children's language/literacy, math, and other learning experiences in the classroom and relating those experiences to their developing skills and competencies.

The purpose of OLOS is to support teachers' efforts to personalize or individualize learning opportunities for each child's unique learning needs so all children in the class are able to achieve greater gains in key outcomes – both academic and social-behavioral. Thus, OLOS focuses on the learning experiences of multiple individual children in the classroom and compares these experiences to what might be more optimal learning opportunities for each child. For example, if a child enters PK with very low language and vocabulary skills, then OLOS can reveal whether they are actually experiencing activities, such as play, book reading and talking with their teacher, that build language – or are they spending most of the time off-task or in activities that do not build vocabulary and language?

There are four principal differences between OLOS and previously developed classroom observation systems: (1) OLOS uses internet technology to allow observers to continuously record the context and content of learning opportunities for multiple children simultaneously and the teacher, while also coding for child and teacher talk (discourse moves) that are embedded in these opportunities. The web-based technology simplifies the coding effort (observers just touch buttons – timing and frequency are recorded automatically) with results available immediately in

the online dashboard; no post-coding or tabulating is required. Please see Figure 1 and 2 for the

OLOS child and teacher coding interface. This differs from other early childhood individual

assessment systems because observers do not have to pause to code or record their observations;

(2) OLOS does not focus primarily on the teacher; (3) nor does it use global ratings of classroom

quality. Rather it is designed to capture differences in the learning environment *within* and

between classrooms by focusing on the experiences of multiple individual children in the

classroom (see also, Connor et al., 2009); and (4) OLOS is designed to be used by practitioners

and is, by design, low-inference. How coding is accomplished is more fully described in the

methods. While some precursors of the OLOS system (described below) have been used in PK

previously (Connor, Morrison, & Slominski, 2006), the entire system has not been tested for

feasibility in PreK. Thus, the purpose of the current project is a feasibility study to examine

possible revisions and adaptations that are needed for OLOS to be used validly and reliably in

PK classrooms.

Developing OLOS

Technology- and internet-based, OLOS observations are conducted on a Chromebook or

other tablet or computer connected to the internet. OLOS builds on and extends earlier

observation systems that focused on children in PK (Connor, Morrison, & Slominski, 2006) and

primary grades (kindergarten – third), and the effect of child-characteristic-by-instruction

interaction effects on children's literacy, mathematics, and science outcomes (e.g., Al Otaiba et

al., 2011; Connor et al., 2006; Connor et al., 2011a; Connor et al., 2011b). Starting with the

premise that children who share the same classroom may have very different learning

opportunities, Connor and colleagues (2009) developed the *Individualizing Student Instruction*

(ISI) observation system to capture amounts of different types of instruction at the level of the individual child. In the ISI observation system, from video, coders record the duration (in seconds) of specific types of instruction (language/literacy, social studies, science, math, or other) as well as the subtypes of instruction (e.g., for language/literacy: comprehension, phonics) that each individual child receives. The instructional context (whole class, small group, or individual) and who is focusing attention on the learning activity (teacher, peers, or child him/herself) are also recorded. These aspects of instruction combine to identify specific learning activities (e.g., a peer-managed, small group, phonics literacy activity). The duration coding portion of the OLOS observation system, recording the amounts and types of instruction, is based on the ISI observation system (Connor et al., 2009). In previous research with the ISI system, the observed amounts and types of instruction, and specifically how different these are from the recommended amounts of instruction based on A2i algorithms, were predictive of children's literacy gains across the school year (Connor, Piasta, et al., 2009).

The child and teacher talk frequency codes of OLOS are based on a second observation system, *COLT* (Connor et al., in press). COLT was developed based on research showing that the discourse environment of the classroom influences child engagement and learning (Almasi, O'Flahavan, & Arya, 2001; Carlisle, Dwyer, & Learned, in press; Murphy, Wilkinson, Soter, Hennessey, & Alexander, 2009). The results of a classroom observation study of 337 children in 50 second and third grade classrooms, using COLT, revealed that the more children and their classmates talked during language/literacy instruction, using nine specific types of talking, the stronger were their reading comprehension gains (Connor et al., in press). Moreover, there were 11 explicit ways that teachers talked to the students that predicted greater use of these nine types

of child talk. While teacher talk predicted child and classmate talk, teacher talk did not directly predict children's reading comprehension gains. That is, child and classmate talk fully mediated the relation between teacher talk and children's reading outcomes. Psychometric analyses of COLT data with elementary-aged children (Connor et al., in press), revealed the child talk codes loaded on to one general factor. This factor predicted child outcomes. In the same way, teacher talk codes all loaded on to one general factor. Preliminary studies with kindergarten and first grade classrooms, in both reading and mathematics, suggest that COLT is also predictive of child outcomes in these grades. However, it is not known whether these types of talk will be observed in PK classrooms. Investigating this is an important aim of this study. We anticipate that greater opportunities for PK children to engage in conversations with their teacher and peers, to think and reason, and to actively participate may be more likely during language/literacy (e.g., shared reading) and math (e.g., calendar time) learning opportunities (e.g., Chien et al., 2010). However, we also acknowledge that whether and how often children talk in class will likely vary across different PK classrooms.

PK classrooms are highly diverse settings that vary in terms of their auspice and sources of funding, the background and experience of teachers, the educational philosophy or beliefs about appropriate instruction guiding the programs, amounts of time devoted to play and exploration, and amounts of time devoted to explicit language/literacy and math instruction (Chaudry & Datta, 2017; Mashburn et al., 2008; Phillips, Gormley, & Lowenstein, 2009). PK classrooms also vary widely in terms of the skills that children bring with them to school (Mashburn et al., 2008). Because of these many differences, it is not clear whether OLOS can be used in different types of PK classroom serving diverse 4-year-olds.

As part of the US Department of Education Institute of Education Sciences (IES) funded

Early Learning Network, the feasibility of adapting OLOS (alpha version) for use in early

childhood programs was tested. We conceptualize feasibility as meeting four criteria: (1)

Research staff are able to be trained and reach reliability on the system in a reasonable amount of

time, (2) Live coders report that they are able to use the system as it was intended based on their

training, (3) Practical and structural barriers do not prevent use in the field, and (4) The system

captures what we intend it to capture (i.e., variability within and between classrooms of

individual children's learning opportunities; teacher and child talk are associated).

The research questions are as follows.

(1) To what extent is OLOS feasible in different types of PK classrooms ($n = 12$

classrooms), which may be less structured learning environments than elementary

school classrooms?

(2) What is the nature of and variability in the amount of time 68 PK children spent in

different learning opportunities (e.g., play, language/literacy, math)?

(3) What is the nature of and variability in the kinds of children's talk and social-

behavioral actions observed in early childhood classrooms and are there multiple

dimensions?

(4) What is the nature of and variability in teacher talk?

(5) Is there an association between teacher talk and child talk?

(6) In what learning opportunity contexts do we observe more and less child talk?

In summary, we aimed to examine how feasibly OLOS can be used in PK classrooms.

Research question one addresses qualitative measures of feasibility such as modifications needed

for PK classrooms, reports of live coders, training experiences, and structural barriers to feasibility when using OLOS to observe classrooms live to address our first three criteria of feasibility. Research questions 2-6 examine the fourth criteria of feasibility; that is, whether we can identify characteristics of learning opportunities (including variability of time in learning opportunities, child talk, and teacher talk) that would be expected if OLOS demonstrates promise of being a useful observation system that captures within- and between-classroom variability in PK children's learning opportunities. We expect to see that children who attend different classrooms experience different learning opportunities and, more importantly, that children within a classroom experience different opportunities to learn in terms of amount and type of instruction and in the talk they use. Additionally, teacher and student talk should be associated.

Method

Child and Teacher Participants

Child participants (n = 68) were from 12 classrooms in four different types of early childhood programs: state PK, state PK– Special Education, Head Start, and tuition-based preschool (see Table 1 for descriptive statistics and number of participants by program). Opt-out consent letters were sent home with all children in each classroom (translated as needed) and at least two weeks were provided for a parent to return the form indicating they did not want their child to participate. All teachers (n=12) provided active consent.

A total of 145 children were invited to participate, three of whom opted out of the study for a total of 142 child participants. One hundred eight of the consented children who met study criteria – 4 years of age – were included in the full year study, and assessed. Of these, 68 were observed in the spring for this feasibility study. Of the children in the study, over 40% came

from low income families, over 25% had an Individualized Education Program (IEP), and about 60% spoke another language besides English at home (see Table 1). Children who attended the tuition-based preschool (PS) were significantly older than the children in the other three programs.

An integral part of OLOS is individual child assessments of vocabulary and literacy (we are adding math in new versions of OLOS), which are computer-assisted and administered individually to students. For this study, the vocabulary and literacy assessments were conducted in the spring prior to the classroom observations. More information on the assessments is available from the first author. The scores on the vocabulary and early literacy measures tended to be higher in the tuition-based PS program, controlling for age, compared to the other programs.

Teacher participants all met state and local qualifications. How many teachers and other adults were in the classroom varied by site. For example, in the state PK classrooms, each classroom had one teacher and one teacher's aide and approximately 15-20 children. In the state PK – Special Education program which served children with special needs as well as typically developing children, there was a maximum of 12 children per class with at least one certified teacher and other adults to keep the ratio 2:1. In the Head Start classrooms, each classroom had one teacher, one teacher's aide and approximately 15-20 children. The tuition-based program was located in a university setting with one master teacher and 4-6 undergraduate trainees. All but the tuition-based program were half day programs.

Early Childhood Programs

Observations were conducted in four types of early childhood programs. Information about these programs was gathered through interviews with the directors and publicly available materials.

State PK. *State PK* is a public preschool in California for children ages three to five that is available free of charge to children whose family income falls at or below 70% of the State Median Income (SMI). The majority of children in the four observed PK classrooms were non-white Hispanic, Spanish-speaking, English language learners. Funded by the California Department of Education, Child Development Division, these PK classrooms were licensed by the State of California Health and Welfare Agency, Department of Social Services. The program sites that we observed were accredited by the National Association for the Education of Young Children (NAEYC). The curriculum was thematic-based and followed the California Preschool Learning Foundations and Framework.

State PK – Special Education. The second program, *State PK – Special Education,* is a public preschool in California serving primarily children with special education needs. The special education program is free to families, and partial funding for the program is provided through preschool grants falling under Individuals with Disabilities Education Act (IDEA) Part B. Children needed to go through the same special education approval process that was expected of all K-12 children in special education. An Individualized Education Program (IEP) was in place for children entering the program. Approximately 140 children were enrolled in this program at the end of the 2017-2018 school year. The curriculum was determined each year

based on assessment results from the Desired Results Developmental Profile (DRDP). No formal observational measures were being used by the school to record classroom quality.

In the current feasibility study, three *State PK – Special Education* classrooms were observed. The types of special needs represented in these classrooms included speech and language disorders, more severe developmental delays such as autism spectrum disorder, and other health impairments including medically fragile children. The demographic backgrounds of these students were different from the other sites: 50% white, 28% Asian, 15% Hispanic, 2% black, and less than 10% of children were English language learners. While state-wide 53% of children are eligible for free and reduced lunch, only 3% at this site were eligible, suggesting that the population was more affluent than is typical for other State PK programs.

Head Start. *Head Start*, a federally funded program, is free of charge for children whose families have an income falling below the federal poverty line. Head Start/Early Head Start programs were funded by the Federal Department of Health and Human Services directly to local community agencies. The majority of these agencies also had contracts with the California Department of Education (CDE), to administer general child care and/or State Preschool programs. In Orange County CA, where these Head Start centers were located, the High Scope curriculum and Numbers Plus for mathematics were used, along with Marvelous Explorations through Science & Stories (MESS). Three out of four children served by Head Start in California were Hispanic and a subset of these were primarily Spanish-speaking. For the current feasibility study, we observed one classroom and teacher at each of four sites (4 classrooms in all).

Tuition-based preschool. The fourth program, *Tuition-based PS*, was entirely tuition-driven but not operating to make a profit. Part of a university, a sliding scale of payment was

available based on family income. Less than 10% of children were English Language Learners. Average enrollment per class was 16 children. This site followed the State Learning Standards and the program identified their own expectations and standards based upon children's strengths and needs. This site used The Creative Curriculum for Preschool (Teaching Stratgies, 2018) and the Hawaii Early Learning Profile (VORT, 2018), which provides a breakdown of skills and milestones by age. They also followed a project or theme-based approach following the children's interests and outcomes they wanted children to achieve. This site had a high volume of research studies taking place and high levels of parent participation. Different combinations of children attended the classroom depending on the day of the week (i.e., some children attended only Tuesdays and Thursdays, others attended just one day a week, while others attended every day).

Standard Observation Protocol

Five trained, reliable coders observed and live-coded classrooms using OLOS in the Spring of 2018. All classrooms were observed on one day for the entire time the children were in the classroom. For half-day classes, the entire morning or the entire afternoon was observed and video-taped. For children who attended full-day classrooms, the entire day was observed and video-taped. Observations were based on the time children were in the classroom. Outdoor activities, library time, and any other activities outside the classroom were not observed.

For each classroom observation, three research assistants were assigned roles: one trained and reliable coder used OLOS to observe live the target children and teacher; a second research assistant videotaped the classroom; while the third research assistant recorded field notes on the

classroom environment and on the children, including recording child descriptions and notes on the classroom organization for later video coding.

The OLOS observer recorded the learning opportunities for three children simultaneously for two cycles. Each cycle in OLOS consisted of 15 minutes observing the same three children simultaneously (see Figure 1) and then five minutes observing the teacher (see Figure 2). Coding was live and in real time; that is the observer was coding throughout the cycles. Figure 3 provides an overview of the procedure. Once the first two 20-minute cycles were completed, three new children were observed for two, 20-minute cycles. Additional children were randomly selected using the procedure described below and observed for as many cycles as the time allowed.

For the feasibility study, observed children were selected randomly by the OLOS technology. Random sampling using the literacy assessments embedded in OLOS is a key feature of the OLOS technology. Before randomly choosing the three students to be observed, a toggle switch allows for filtering of students who are below grade expectations or at/above grade expectations, and then OLOS randomly selects from among this filtered list of children when the user presses the "Select Random Student" button (see Figure 10). OLOS also allows observers to intentionally select a particular child. For this study, we used the random selection feature of OLOS, selecting two children who were below age/grade expectations and one at or above expectations. This was because we aimed to determine whether OLOS could distinguish between children with higher and lower literacy skills. The system functionality that allows for randomly choosing children based on whether they are performing below or at grade level has the potential

to empower practitioners to determine if there are systematic differences in the classroom experiences of children with different achievement levels.

Duration of Learning Opportunities

Operating OLOS on touchscreen tablets, observers simultaneously recorded the following learning contexts for the three individual children being observed: child behavior; content of instruction; subtype of instruction; and context of instruction. We describe each below.

Child behavior. Child behavior recorded whether the child was on- or off-task. *On task* was defined as being engaged in the learning activity while *off task* was coded when the child was visibly disengaged from the learning activi**ty.**

The content of instruction. The content recorded included language/literacy, math, science, social studies, art, music, other, and non-instruction. *Language/literacy* referred to activities related to language development, decoding, comprehension, fluency etc. *Math* referred to activities related to the understanding and manipulation of numbers. *Science* was coded when the point of the lesson was to teach science (not reading) and content was related to topics such as life cycles, identifying parts of plants or animals, or basics of how to conduct an experiment, etc. *Social studies* was coded when the point of the lesson was to teach social studies (not reading) and the content was related to topics such as government, geography, or history. *Art* was coded when the activity focused on art for art's sake and was not a literacy-related or math-related art project. Similarly, *music* was coded when the activity involved music for music's sake and was not related to any other content area. *Other* was chosen as the content code when children were being taught lessons content that was educational but did not fit in any of the other defined content areas – lessons on behavior and social skills were coded as *other*. Non-

instructional time was also an available option under the content heading. *Non-instruction time* includes non-academic time (e.g., nap, brushing teeth), as well as transition between activities, disruptions, discipline, or waiting for the teacher (Day & Connor, 2016). In this version of OLOS, *meal time* was coded as non-instruction time.

The subtype of instruction. When observing language/literacy or math, OLOS allowed for recording the subtype of the learning opportunity. The subtypes of language/literacy content are code- or meaning-focused instruction and play. *Code-focused instruction* referred to language/literacy activities related to identifying letters, letter-sound associations, and writing letters (with more advanced skills, such as spelling and phonics for upper grades) whereas *meaning-focused instruction* referred to language and literacy activities that focused on vocabulary, shared book reading, and other activities designed to build language, comprehension, and writing composition skills. We anticipated that the latter would be observed in upper grades rather than PK. *Play* referred more generally to time in early education classrooms that are most frequently opportunities to build language and literacy skills and included choice (e.g., dramatic play centers). Math subtypes are number knowledge, operations, geometry, algebra, and applied math. *Number knowledge* referred to activities that required children to recognize numbers, count, or, for older children, understand decimals or fractions. *Operations* consisted of addition, subtraction, multiplication, and division activities. *Geometry* was coded when the activity focused on shapes, angles, lines, or transformations. *Algebra* consisted of pattern recognition and solving for unknowns. *Applied math* was coded when the activity related to calendar, time, money, measurement, data analysis, or probability.

The context of instruction and management. Using OLOS, the context of instruction (i.e., whether the child was learning in the context of a small group, the whole class, or individually) was coded. *Small groups* consisted of between two to eight children. Any groups larger than eight children were coded as *whole class*. Moreover, the management of these contexts (the person focusing the attention towards the instruction) was also coded. This allowed us to record whether the child was with the teacher (teacher-child-managed) with peers (peer managed), or working alone.

For each child, the time spent in particular combinations of contents and contexts was recorded continuously throughout the observation. For example, during the observations of the three children selected for the 15-minute observation cycle, one child might be working directly with the teacher on a letters activity and would be coded as follows: Child behavior: On task; Content: Language/Literacy; Instruction subtype: Code-Focused; Context: Individual; Management: with Teacher.

A second child might be working in a small group with her peers to count and sort colored blocks and would be coded as follows: Child behavior: On task; Content: Math; Instruction subtype: Number knowledge; Context: Small group; Management: with Peers. The third child could be on task; Content: Science; Instruction subtype: N/A; Context: Individual; Management: Alone. Child behavior, context, and management are coded across content areas, but instruction subtypes were coded only for language/literacy and math content areas. All of these child learning opportunity context codes were initially developed and validated in the ISI observation system (Connor et al., 2009) and the content and context codes are combined to describe learning activities.

Frequency of Classroom Instructional Discourse

In addition to coding the time spent in different learning opportunities, the OLOS observer recorded instances of child and teacher talk during these learning opportunities using the validated COLT codes (Connor et al., in press). Tapping the code on the Chromebook screen records one instance of either child or teacher talk (see Figures 1 & 2).

Child talk. There were nine types of child talk (discourse moves) that were recorded. The first three (highlighted yellow in Figure 1) are as follows. *Non-verbal response* is used in moments when children use their body to respond or attempt to respond to a question related to instruction (e.g., raising their hand to answer a question; standing up in response to teacher prompt to "stand up if you think the answer is 10".) *Verbal response to question or prompt* captures moments when children respond to questions or prompts related to instruction with simple responses that do not require inferencing or predicting (e.g., what was the main character's name?). *Reading text/problems aloud* is used when children are reading text or math problems out loud. In PreK, children were given credit for this code when reading single letters or numbers aloud.

The next five codes (highlighted in green in Figure 1) are used to track higher-level child participation. *Answering questions that require reasoning* captures moments when children answer questions that are not just explicit recall of facts but require some higher-level thinking. *Using text to justify reasoning* is used when children provide evidence from a text or word problem to support their answer. *Participating in a discussion* captures times when at least three exchanges occur, each building on a single topic. *Voicing a disagreement* occurs when children state a disagreement about an instruction-related topic. These seven codes were taken directly

from the COLT coding system and have been shown to be related to child reading

comprehension outcomes (Connor et al., in press). Again, we did not anticipate that all of these

codes would be observed in PK classrooms.

Two additional child frequency codes were added on the recommendation of school

partners. The predictive validity of these two codes has not yet been established. The *uses words

to resolve a difficult social situation* and *moves away from difficult social situation* codes were

intended to capture moments when children are independently displaying good self-regulation of

behavior/emotions in the classroom.

Teacher talk. The teacher frequency codes, nine are from COLT, capture the frequency

of teacher's talk that is intended to support children's active participation in learning

opportunities. Tapping the code on the Chromebook screen records one instance of teacher talk.

The first three codes (highlighted in red in Figure 2) are conceptualized as promoting general

child participation. *Invites children to share information* is used when the teachers invite the

participation of more than one child either in answering a question or otherwise participating

verbally in a task. *Summarizes or synthesizes child responses* is used either when the teacher

rewords a single child's or multiple children's answers to make it more comprehensible to the

class or when the teacher synthesizes responses from two or more children. *Expresses interest in

child responses* captures moments when the teacher comments specifically on what they liked

about a child response or otherwise markedly makes a child feel positively about an answer

he/she provided; simply noting "good response" is not sufficient.

The second two codes (highlighted in yellow in Figure 2) capture moments when the

teacher is teaching content. *Explains topic-specific concepts* refers to moments when the teacher

teaches a lesson about language/literacy or math concepts that are relevant to multiple lessons, such as identifying the main idea, learning how to scan for key words in compare and contrast texts, or identifying rectangles. *Provides background information* is used when the teacher provides children with specific information that will help them understand a single lesson or text (e.g., discussing what a "snow" is before reading A Snowy Day -- the feasibility study was conducted in a warm climate; talking about what a dollar is before children play in a store dramatic play area).

The next four codes (highlighted in green in Figure 2) are designed to record teacher support for higher-level child participation. *Directs children to use evidence and explanation to support answers* captures moments when the teacher prompts a child to use specific information from a text or to generate an example to justify a response. *Challenges children to reason or draw conclusion* is used when the teacher asks questions that require higher-order thinking on behalf of the child such as inferencing or predicting. *Asks follow-up questions to gain additional information or clarify an idea* is given when the teacher asks a single child a follow-up question related to instruction. *Encouraging children to make connections to self or other text or problem* captures times when the teacher prompts children to think about how a lesson is related to their own lives or to another text or concept they already know.

Finally, the last four teacher codes are those that were adapted from the Quality of the Classroom Environment (Q-CLE; Connor et al., 2014) observation system to count instances of teacher talk that supports child social and behavioral outcomes. Essentially, we translated the Q-CLE Likert scale to observable behaviors. *Handles disruptions quickly and efficiently* is used when teachers are able to avoid interrupting instruction for extended period of time when a

disruption occurs in the classroom. *Redirects child misbehavior in a positive way* is used when teachers quickly redirect children without interrupting instruction and usually by pointing out the correct behavior instead of drawing attention to the incorrect behavior. *Uses a positive behavior management strategy* is used each time the teacher uses an observable strategy (e.g., point systems, color-coded groups) that allows her to more efficiently manage behavior in the classroom. *Smooth and orderly transition between classroom activities* is used when the teacher is able to support her children in moving quickly and quietly from one activity to the next. Scaling and predictive validity have not yet been determined for these codes.

Training to Conduct OLOS Observations

Observers were undergraduate and doctoral-level students in Education Sciences trained to use OLOS by participating in an in-person training of approximately 10 hours in length occurring over three separate sessions. Observers then practiced coding using previously-recorded videos of classroom observations, and then tested their reliability by coding a reliability test video. The bank of reliability practice and test videos consisted of 10-minute clips of previously recorded classroom observations of both math and language/literacy instruction cut from longer (60-180 minute) recordings. Coders first practiced coding students and teachers by coding videos until they exceeded 70% agreement with the gold standard coder on two practice videos, then tested reliability by coding a reliability test video once. If the observer exceeded 70% reliability with the gold standard coder on the test video, they were considered reliable. If they failed the reliability test, they coded new practice videos until reaching the reliability threshold and then took a new reliability test. This cycle continued until the coder became reliable or failed five consecutive reliability tests, in which case it was considered that the coder

was not able to become a reliable user of OLOS. All coders were able to become reliable with a maximum of four reliability tests.

Based on coder reports (n = 6), it took an average of four to five hours of video coding and two reliability tests after completing the training to become a reliable user of OLOS. Using Noldus Observer Pro, inter-rater agreement (Kappa) was calculated for observations of three students coded simultaneously as well as for the teacher observation. Coders were required to achieve a minimum Kappa of 0.70 with gold standard videos. Interrater agreement ranged from 0.78 to 0.84.

To establish live coding reliability, two of the trained observers observed the same children in the same classrooms at the same time, 12 children in all. We calculated inter-rater agreement for the duration and number of activities children experienced, content of the activities (e.g., language/literacy, math), and for child talk frequency codes (e.g., ask questions, verbal response to question) collectively and individually. As a minimally viable product, duration codes were more difficult to analyze than they are with the current version. For this study, to calculate inter-rater agreement for duration, we exported the duration file then created a total-seconds observed in each content area for each observer. So, for example, Coder 1 observed a total of 9706 seconds of non-instruction and Coder 2 observed a total of 6109 seconds of non-instruction. We then divided the smaller by the larger number and multiplied by 100 to get a percentage of agreement (in this example about 63%). We repeated this process for each content area. Overall, the reliability was approximately 62%. We then counted the number of different learning activities (language/literacy, math, art) each coder observed and then divided the smaller number by the larger number and multiplied by 100. Results revealed 74% agreement on the

number of different learning activities observed. We next compared agreement on the content of

these activities using the smaller number observed and found 77% agreement on the content of

these activities.

We calculated inter-rater reliability for the child talk frequency codes by summing the

number of codes where the observers agreed and dividing by the total number of codes. Overall,

inter-rater agreement was 96% for the child talk frequency codes overall. Considering the child

talk codes individually, inter-rater agreement was above 87% for all codes (e.g., *non-verbal*

responding = 87%; *verbal response to question* = 89*%; reading text/problems* aloud= 91%). For

the social-behavioral codes, inter-rater agreement was 94% for *in a challenging social situation,*

uses words to resolve issue and 98% for *moves away from difficult social situations*.

Analytic Strategies

Research question 1. To qualitatively examine the feasibility of OLOS for PK

classrooms, we utilized observer notes and interviews. The observers kept field notes, which we

repeatedly examined for evidence of systematic issues with the technology or with the user

interface . We used this information to help evaluate and improve the feasibility of using OLOS

live in PK classrooms after discussing the findings from the field notes with the observers. We

also relied on informal and formal interviews with school partners including teachers and

administrators. Finally, we examined the reliability scores and the reported time it took each

observer to reach reliability.

Research question 2. To determine whether OLOS was able to capture within- (and

between-) classroom variability as intended and to examine the nature of and variability in the

amount of time PK children spent in different learning opportunities, the duration data for

children were downloaded directly from the OLOS system. OLOS calculates the duration in seconds of learning opportunities. To determine the nature and variability of learning opportunities we ran descriptive statistics in SPSS (version 24).

Research question 3. To examine the nature and variability in the kinds of children's talk and social-behavioral actions and whether there are multiple dimensions, we utilized the raw frequency of the child and teacher talk codes per learning opportunity. The datasets, including the frequency codes, were provided by the OLOS system in Excel files. Similar to the previous question, we ran descriptive statistics in SPSS to determine the nature and variability in the kinds of children's talk and social-behavioral actions. Also, we report intraclass correlations representing the variability in child talk that falls between classrooms, which was calculated using the Hierarchical Linear Modeling statistical software (HLM, version 7; Raudenbush, & Bryk, 2002). To determine the dimensionality of the child talk codes, we used exploratory factor analyses also in SPSS.

Research question 4. Similar to the previous research questions, to examine the nature and variability in teacher talk, we utilized the raw frequency dataset downloaded from OLOS and ran descriptive statistics in SPSS.

Research question 5. To examine the association between teacher and child talk hierarchical linear modeling (HLM) was used, because children are nested in classrooms. Using HLM prevents the mis-estimation of standard errors. To obtain the zero-order association between teacher and child talk, we used child talk as the outcome, removed the intercept, and then added teacher talk to the model at the classroom level (see Equation 1), where child talk for

child i in classroom j is a function of the effect of teacher talk, γ_{00}, with classroom and child level

variance, u_{0j} and r_{ij} respectively.

$$Child\ Talk_{ij} = \gamma_{00}*Teacher\ Talk_j + u_{0j} + r_{ij} \qquad\qquad (1)$$

Research question 6. To determine the learning contexts in which more student and

teacher talk occurred, we used zero-order correlations generated by SPSS.

Results

Research Question 1: Feasibility of OLOS in PK Classrooms

In this section, we report on four sources of information that allowed us to evaluate

feasibility: (1) amount of time it took to become trained and reliable on the system, (2) coding

modifications needed specifically for PreK classrooms, (3) reports from live coders, and (4)

structural barriers to implementation in the field.

After completing a 10-hour, lecture-style training, all coders were able to reach the 70%

threshold of reliability in coding both duration and frequency codes when coding students and

teacher after an average of two reliability tests (minimum=1, maximum=4) and approximately 5

hours of video practice coding. Coders were required to achieve a minimum Kappa of 0.70 with

gold standard videos; Kappas achieved ranged from 0.78 to 0.84.

Not surprisingly, there were challenges in adapting the elementary school version of

OLOS for PK classrooms. For example, a number of the child talk codes focused on reading text

or reading math problems, which had to be re-defined to be appropriate for PK classrooms (e.g.,

single letters and numbers counted as text). Other child talk codes, such as "using text to justify a

response" were never observed. There was a consensus that more talk codes related to social-

behavioral development were needed. One observer noted, "The frequency codes don't capture a

lot of student and teacher talk that could improve socio-emotional learning." Another reported, "I

would add/change some of the socio-emotional frequency codes, at least on the teacher's end

(e.g., models appropriate behavior, prompts students of expected behavior). I'd like to see socio-

emotional codes that we think can be predictive of some aspect of self-regulation or self-

regulation's effect on academic growth or instructional quality." New socio-emotional frequency

codes are currently in the development and testing phase for future inclusion in the OLOS

system.

For the content-context duration codes (e.g., language/literacy, math, play), unlike many

early elementary classrooms, there were not blocks of time devoted to language/literacy and

math. Rather these activities frequently occurred simultaneously (e.g., there are both library

corners and counting play centers). Because individual children were often involved in different

types of activities at the same time, we re-designed OLOS to accommodate a range of content

areas simultaneously. One observer recorded:

> Other concepts aside from literacy seem to look different in Pre-K classrooms
>
> than with the older kids. For example, I have encountered a scenario where they
>
> are looking at a measuring cup filled with water and ice and discussing whether
>
> the water will go up or down if the ice melts, and when you are live coding it is
>
> hard to decide whether it should be math because they are talking about volume or
>
> science because they are discussing the different types of matter.

Another difference in PreK classrooms was that there tended to be a much larger

proportion of time spent in play compared to Kindergarten through 3rd grade classrooms. One

observer reported that "I feel like play should be its own instructional content category, separate

from a subtype of literacy." In this version of OLOS, play is coded as a language/literacy

activity. Observers suggested that by making play at the same level as language/literacy or math,

more detail on the type of play could be captured.

Additionally, observers noted that it was challenging to observe children in different parts

of the classroom. One child might be engaged in a language/literacy learning opportunity while a

classmate might be involved in a science learning opportunity, and then switch learning contexts

only 5 minutes later. Additionally, teachers tended to move around the classroom and interact

with different children briefly, which made coding whether the student was with the teacher or

with peers challenging. Overall, observers noted that context and content change more rapidly in

PK classrooms compared to elementary classrooms. Nevertheless, observers reported that coding

three children at a time, but no more, while maintaining accurate coding of content, context and

type of instruction as well as student and teacher talk was feasible. One observer reported,

"Finding the right place to stand in the PK classrooms is important because you need to be able

to track what 3 different kids are doing and how they move from one activity to the next. I could

do this with three students, but not any more than that".

There were some structural and logistical barriers to feasibility in the field. Limited or

unstable internet connection resulted in difficulties for the end user (i.e., error screens) as well as

abnormalities in the data. We used wireless hotspots instead of the school internet when

necessary to avoid these issues. Data were cleaned to correct errors identified related to unstable

internet connection. Additionally, there were some issues related to the web-based technology

(e.g., users being automatically logged out during breaks longer than 20 minutes) and other

technology limitations including user error (unintentional button presses) when using the OLOS

prototype (i.e., alpha testing). During the school year we made iterative changes to the OLOS programming code (to correct any system errors identified) and to the user-interface (e.g. made frequency code buttons larger and "finger-shaped") based on feedback from school partners and the observers in the field. We are currently working on a version of OLOS that does not rely as heavily on the wireless connection of our school partners.

Research Question 2: What is the nature of and variability in the amount of time PK children spend in different learning opportunities (e.g., play, language/literacy, math).

In this section, we provide descriptive statistics about the extent to which the current frequency and duration codes included in OLOS were observed in PK classrooms. We also consider this a measure of feasibility inasmuch as we are tracking whether the content areas and types of student and teacher talk that were included in OLOS are relevant to and actually occur in PK. We then discuss duration, frequency, and between and within classroom variability.

The student and teacher talk codes were examined to determine how frequently they occurred (See Figures 8 and 9 for mean frequency per observation and total frequency per teacher). All but one child talk code was observed at least five times across the observations, with non-verbal and verbal responding observed the most (196 and 248 times, respectively), followed by reading text aloud (63 times). The higher-level codes such as answering questions that require reasoning and participating in a discussion were, perhaps unsurprisingly, observed less frequently – an average of 12.2 times across all observations. Finally, the socio-emotional codes of using words to resolve a difficult social situation and moving away from a difficult social situation were observed quite infrequently (just 9 and 5 times, respectively). All but one of the teacher talk codes were observed.

In terms of the duration of content area instruction, all areas were observed with the exception of Social Studies. We refer the reader to Table 2 and note that language/literacy and non-instruction were the most commonly observed instruction types, followed by Art and Play. Math and science were observed very infrequently in PK.

For this research question, we focused on the content of the learning opportunities afforded to each child in the class. Thus, a total duration (in minutes) for each content area (i.e., language and literacy, mathematics, science, social studies, art, music, play, and other) was computed for each child standardized on 30 minutes (duration in min divided by the total observation time in min*30). Children who were observed less than one full 15-min cycle were not included in the analyses.

On average, children spent much of their time in non-instructional activities (e.g., lining up, waiting for the teacher, meal time, see Table 2 and Figure 4) – almost one-third of the 30-minute observation. However, this varied substantially for individual children ranging from 0 minutes to 28 minutes. Also notable was that in this iteration of OLOS, meal time (e.g., lunch, snack) was considered non-instructional. Because we observed that meal time offered opportunities for rich talk and discussion, the new version of OLOS records meal time separately from non-instruction.

Many children spent substantial amounts of time in *language/literacy activities* (see Table 2); about 9 minutes (of the 30-minute observation period) on average, and this ranged from 0 to more than 20 minutes. About 5 minutes, on average, was spent in *play* activities and about 4 minutes were spent on *art* and *music*. In general, children spent very little time in math and

science although this varied by child ranging from 0 to 12 minutes. No social studies learning opportunities were observed.

To provide a sense of how variable learning opportunities were for children within and between classrooms, we provide the minutes of language/literacy observed for each child in Figure 5. There was notable within classroom variability. We see that most of the children in all of the classrooms and programs participated in language/literacy learning opportunities. In contrast, only ten children participated in science learning activities (see Figure 6).

When we considered children's *off-task behavior*, keeping in mind that we were still recording the content of the instruction they were supposed to be receiving, we found substantial variability among children with time off-task ranging from 0 to 30 minutes. However, the mean time was quite low; just under 2 minutes on average suggesting that children were generally on task.

Research Question 3: What is the nature of and variability in the kinds of children's talk and social-behavioral actions observed in classrooms?

The HLM intraclass correlation (ICC) for total child talk frequency was 0.387, with children nested in classrooms, indicating that about 39% of the variability in child talk frequency was explained by the classroom they attended (i.e., between classroom variance).

In previous studies of child discourse moves among second and third graders, the child talk codes loaded on one general dimension. To examine whether this was the case for younger children, we conducted a preliminary exploratory factor analysis using generalized least squares in SPSS. Our results suggested that the child talk codes loaded on four overlapping factors with one strong general factor (see Appendix Table A.1). All the types of talk, except the social-

behavioral talk, loaded on the strong general factor. For the second factor, nonverbal responses

loaded highly. Notably, asking questions negatively loaded on this factor. For the third factor,

answering questions loaded negatively and strongly on this factor. The fourth factor included

only the social-behavioral codes. The four factors together explained 60% of the variance with

the first factor explaining about 23%. Using these results, we created a child talk variable that

summed all the kinds of talk observed in 30 min, except *reading letters and numbers* (i.e., we

included variables with factor loadings > .35).

Overall, in 30 minutes, children were observed to use between 6 and 7 discourse moves

but this ranged from 0 to 52 instances of child talk (see Table 2). Figure 7 shows the amount of

talk for each individual child within the 30 min observation. Most of the children talked at least

once and thirteen talked 10 or more times. Again, there was notable within and between

classroom variability.

There were fewer instances of children demonstrating social-behavioral skills by moving

away from challenging situations or using their words to resolve conflicts. On average, we

observed these actions between 0-1 time (mean = 0.52 per 30 min) but again, the range was

great; from 0 to 21 times per 30 min.

Research Question 4: What is the nature of and variability in teacher talk?

An examination of talk frequencies revealed that teachers used all kinds of talk, with the

exception of *encouraging students to use evidence to support their answers* (see Figure 8 top).

Of the teacher social-behavioral support actions, *redirecting behavior* was the most frequently

observed (see Figure 8, bottom). Because there were only 12 teachers, we did not conduct a

factor analysis. Rather, we created total frequency variables by summing the frequencies of all

observed teacher talk and all observed teacher social-behavioral actions. Total talk and social-behavioral actions for individual classrooms is provided in Figure 9. The variability between classrooms is notable.

Research Question 5: Is there an association between teacher talk and child talk?

Based on previous evidence from related observation systems used in elementary school classroom, we expected an association between teacher and child talk. Thus, we examined whether there was an association between child talk and teacher talk using HLM with children nested in classrooms. We found that teacher talk frequency was positively and significantly associated with child talk frequency (unstandardized coefficient = 0.272, $t(12) = 0.108$, $p = 0.027$). Thus, for every increase of 10 instances of teacher talk, children's talk frequency would be expected to increase by 2.7 instances. There was significant classroom level variance (var = 7.99, $X^2(12) = 77.30$, p<.001) and child level variance (var = 6.087). There were no significant associations with teacher social-behavioral frequency score and child talk frequency or self-regulation moves. Of note, we did not consider child differences by site.

Research Question 6: In what learning opportunity contexts do we observe more and less child talk?

To determine whether OLOS is successful in describing PK classroom talk, we should see differences in child talk based on context of learning opportunities. Using zero-order correlations, we examined associations among child talk, social-behavioral actions, and different learning opportunities (e.g., language/literacy, play, math). Results indicated that child talk was observed more frequently during language/literacy activities ($r = .337$, $p = .005$), during math ($r = .266$, $p = .028$), and during science learning opportunities ($r=.501$, $p<.001$). Less child talk was

observed during non-instruction (r = -.275, p =.023). The only context in which we were likely

to observe children's positive social-behavioral actions was during non-instruction (r = .341, p

= .004). Children were less likely to be off task during language/literacy instruction (r = -.373, p

= .002).

Discussion

The results of this OLOS feasibility study, using four criteria, revealed several important

findings. Again, the four criteria were: (1) Research staff are able to be trained and reach

reliability on the system in a reasonable amount of time, (2) Live coders report that they are able

to use the system as it was intended based on their training, (3) Practical and structural barriers

do not prevent use in the field, and (4) The system captures what we intend it to capture (i.e.,

variability within and between classrooms of individual children's learning opportunities;

associations between teacher and child talk). Researchers were able to train and become reliable

on OLOS in a reasonable period of time. Live observers reported that implementation in the field

was feasible for live coding of up to three students simultaneously. Structural, technological, and

practical barriers to implementation in the field did exist, but this study allowed for the

development of changes to the system and strategies to resolve these issues. Finally, OLOS

captured differences in children's learning opportunities both within and between classrooms and

the data with this sample of 68 children in 12 classrooms were revealing.

OLOS is composed of observation systems that had been developed and validated for

children in early elementary classrooms; specifically the ISI system, which captures the amount,

content, and context of instruction, and the COLT system, which captures child and teacher talk

(Connor et al., in press; Connor, Morrison, et al., 2009). But it was not clear that OLOS would be

feasible in PK classrooms. This feasibility study revealed that observers were able to gain

acceptable reliability in a reasonable amount of time and could use OLOS to capture learning

opportunities in a variety of PK classrooms that differed in funding streams and curricula. We

also examined whether OLOS in PK provided results that were similar to findings in early

elementary classrooms although we expected some differences. We did find that children who

shared the same classroom had different learning opportunities (within classroom variability).

These learning opportunities varied between classrooms as well. The amount of child talk

observed for individual children also varied within and between classrooms and, importantly, by

context (e.g., literacy, play). The amount of teacher talk and social-behavioral actions varied

substantially by classroom. Finally, congruent with previous findings, the more teachers talked in

specific ways, the more likely the children in their classroom were to talk.

As described in the introduction, OLOS was developed from three integrated, valid and

reliable coding systems: the ISI coding system (Connor, Morrison, et al., 2009), which

documents the amount of time children spend in different learning activities; the COLT system

(Connor et al., in press), which captures nine child and 11 teacher types of discourse moves (i.e.,

talk) that are associated with child learning; and the Q-CLE (Connor, Morrison, et al., 2009),

which provides a global rubric of teacher warmth and responsiveness, discipline, and classroom

organization. With the exception of the ISI coding system that was used in PK classrooms, the

other two systems were validated only in early elementary classrooms. Thus, it was encouraging

that using OLOS was feasible when observing PK classrooms, which differ substantially from

the generally more structured elementary classroom setting.

We intentionally recruited classrooms from four different types of PK programs – Head

Start, State PK, State Special Education PK, and Tuition-based PK – to determine feasibility in

these different settings. OLOS was feasible in these very different programs. At the same time,

the observers noted that there were challenges to using OLOS in PK classrooms that some

adaptations would alleviate. For example, it was challenging to make decisions about the content

of instruction (e.g., math or science) in the more fluid and changing learning environment of the

PK classroom. To address this, we clarified and provided examples in the coding manual and

during training. For sharing results with teachers and coaches, subsequent to this study, we

developed a teacher dashboard so that observation results are immediately displayed at the end of

the observation in easy-to-read charts and tables. We also added a code "social-emotional

curriculum" to the *other* category so that we could more closely monitor explicit learning

activities focused on social-emotional learning.

As hoped, OLOS captured different learning opportunities for children who shared the

same classroom. For example, in Figure 5, which shows the amount of time individual children

spent in language/literacy activities, within-classroom differences in the amount of time in

language/literacy activities were apparent. This is the case even though we always observed three

children at the same time. Even in classrooms where children generally talked a lot, there were

still substantial differences among the children in the frequency of their talk.

We argue that this within classroom variability is important to capture and that by

observing individual children, we can gain insights into the classroom learning environment that

are not available with more global classroom-level measures. Previous research suggests that

classroom quality may have differential effects on children's language and learning outcomes in

the same classroom based on factors such as ethnicity and socioeconomic status (e.g., Burchinal, Peisner-Feinberg, Bryant, & Clifford, 2000; Garcia Coll, 1990; Lamb, 1998; McCartney, Dearing, Taylor, & Bub, 2007). Therefore, by observing multiple individual children and documenting their learning experiences, we can gain insights into the classroom learning environment as it impacts each child. This is not currently possible with more global classroom measures or observation systems that observe one child.

OLOS also captured substantial classroom differences in the extent to which children talked in ways that are generally associated with stronger learning (Connor et al., in press). Again, within and between classroom differences were evident (see Figure 7). Preliminary exploratory factor analyses suggested some dimensionality to the types of child talk with one strong factor for child talk and another for the social-behavioral actions. Further investigation with a larger sample is needed. Not surprisingly, since OLOS is being designed to be used in PK through third grade classrooms, some of the child and teacher talk codes were not observed in the PK classrooms. These tended to be higher level discourse moves, such as using text to justify a response (child talk) or asking students to use evidence to justify their answer (teacher talk). Because the intention is for OLOS to be used by practitioners from PK through third grade, we expected some codes to be infrequent in PK classrooms but to be more frequent in the later grades. At the same time, we did observe children using sophisticated discourse moves, such as answering questions that require thinking and reasoning.

Importantly, as we have found in our previous studies with COLT, which provided the talk codes, more teacher talk predicted more child talk (Connor et al., in press). This is an encouraging finding inasmuch as child talk may be encouraged by specific discourse moves

teachers make, with the potential to improve the classroom learning environment; how teachers interact with the children in their classrooms has implications for children's active engagement and learning (Downer et al., 2010). It was beyond the scope of this feasibility study to examine whether child and teacher talk was associated with child learning outcomes, but other research indicates that the kinds of talk that are captured by OLOS are associated with stronger child outcomes in both literacy and mathematics (Connor et al., in press; Hill et al., 2008; Murphy et al., 2009; Snow, 1989).

The context of learning activities was associated with differing amounts of child talk. Children were more likely to use the coded types of talk during more academically-oriented activities – language/literacy, mathematics, and science. Children were also less likely to be off task during language/literacy activities. This suggests that young children are actively engaged and participating in more academic kinds of learning opportunities, which is encouraging given the accumulating research about the importance of children's early literacy and mathematics development for their later academic success (Duncan et al., 2007). Not surprisingly, less academic talk was observed during non-instructional activities (i.e., waiting in line, transitions). However, children were more likely to be observed using positive social-behavioral actions during non-instruction. It may be that during such unstructured time, there are more opportunities for children to need to use these actions. However, this is conjecture and the social-behavioral actions were only infrequently observed. More research is needed.

While the results of the present study are preliminary, they demonstrate the potential OLOS has to provide useful information to practitioners about the types of learning opportunities their students are receiving and also how these children are participating within those learning

opportunities. By capturing such information through continuous, live observations of the classroom, OLOS stands apart from other snapshot coding systems that require brief observations with pauses for the coding of the observation. However, it is important to note that individual children were coded for approximately 30 minutes each during the live observations and, therefore, some of the variability in learning opportunities observed was random; that is, simply due to what was happening in the classroom when a specific child happened to be selected to be observed. Video coding will be completed that codes all target children for the entirety of the observation to determine how the 30-minute observations compare to full-day observations.

Additionally, the aim of OLOS is *not* to be a high-stakes evaluation of teachers' performance; rather it focuses on children and is designed to be a flexible tool that practitioners can reliably use to provide teachers with detailed descriptive information about the learning experiences of their students and to promote the types of learning opportunities and teacher instructional discourse moves that are demonstrably related to child outcomes. As noted in the introduction, the purpose of OLOS is to support teachers and coaches in individualizing or personalizing the instruction children receive. Thus, it is important that OLOS is sensitive enough to elucidate that children who share a classroom experience different learning opportunities. Teachers and coaches can then better match children with the learning activities that will be most effective for them based on their constellation of skills (e.g., Al Otaiba et al., 2011; Connor et al., 2006). The new teacher dashboard was designed to provide this information to teachers and coaches.

Although the classroom sample size here was too small to draw conclusions about differences between program types, eventually with increased numbers of each type of classroom, the data should allow for quantitative comparisons of differences in learning opportunities in different program types with different curriculum and funding streams. This has clear policy implications for how to fund and design early childhood programs based on the aims of the program. For example, we observed relatively more literacy and math instruction in the state-funded PK, whose explicit aim was building early academic skills. Policy makers may also be interested in results of observation systems that practitioners can use, are low stakes, and that are more child-focused, such as OLOS, which may provide more nuanced indicators of the classroom learning environment.

There are limitations to the study that should be considered when interpreting these results. First, in this study, we used data from observations conducted in early spring. Classroom learning environments, particularly for young children, can change greatly from fall to spring. We might have found very different results had we observed at different times and observed different children since children were randomly selected for observation. There is some evidence that adult talk is fairly stable over time and across observations (Bowers & Vasilyeva, 2010; Huttenlocher, Vasilyeva, Waterfall, Vevea, & Hedges, 2007; Nelson & Bonvillian, 1973; Smolak & Weinraub, 1983). Additionally, there were child differences between sites that we did not consider in these analyses. Finally, these results should be interpreted in light of the policies and funding streams for California and the types of PK programs included in this study. They may not translate to other parts of the US and the world or to other program types.

The data presented in this paper represent the first live use of the OLOS technology in early childhood classrooms as part of the design and development studies we are conducting to develop OLOS as a viable observation tool that can be used by practitioners, that provides instant results, and that captures the nature of and variability in learning opportunities provided to multiple individual children from PK through third grade. We are pleased with the progress so far and find these results promising. OLOS users achieved a high level of inter-rater reliability on both duration and frequency coding in a relatively short period of time. It is important to note that the feasibility study reported here refers to feasibility of use by the research team, and not by practitioners. Currently, we are working with seven school-based literacy coaches to determine the feasibility of use by practitioners. These coaches were able to learn OLOS and achieve high agreement with gold standard videos before the end of a three-day workshop. How they use OLOS with their teachers is being studied.

This study also elucidated design changes that will make OLOS a better tool. For example, in the iteration of the observation tool described in this paper, meal time was coded as non-instruction. However, meal time can be a good environment for building children's language and social skills. Thus, we have now added meal time as a specific learning activity at the same level as art and science. In addition, it became clear that not all coders and practitioners were aware that oral language formed part of the literacy content area, so we renamed the literacy content area "language/literacy", as it has been referred to in this manuscript, to make this more explicit.

It was beyond the scope of the current analyses to examine whether or not children's learning opportunities, captured through OLOS, predict their language/literacy, mathematics, and

social-behavioral outcomes. We are conducting a nation-wide predictive validity study over the

coming years that will investigate the extent to which OLOS captures child learning

opportunities that are associated with gains in early literacy, mathematics, and socio-behavioral

outcomes. One of the previous challenges to exploring differential effects of classroom quality

on individual children has been limitations in power due to the cost associated with recruiting

participants and administering in-depth assessments at multiple time points (Burchinal et al.,

2000; McClelland & Judd, 1993). The predictive validity study with OLOS has the potential to

overcome this barrier thanks to the partnerships created as part of the IES-funded Early Learning

Network. Data are being collected with multiple children in at least 28 classrooms in six different

states (at least 168 classrooms) allowing for a diverse, representative sample of early childhood

educational contexts in the United States. Once these data are collected, all consented children

and the teacher will be coded from video for the entire observation, allowing us to look at how

time spent in different learning environments, and how well these opportunities match with

children's constellations of skills to affect their growth in literacy, math, and socio-behavioral

skills across the school year. In addition, we will investigate how student talk and teacher talk

and their interaction with time in instruction impact student's growth in these areas. Our aim is

that, broadly implemented, OLOS can become a powerful technology tool for teachers and other

practitioners in understanding how to improve learning environments and opportunities for all

children from PK through third grade.

References

Al Otaiba, S., Connor, C. M., Folsom, J. S., Greulich, L., Meadows, J., & Li, Z. (2011).
Assessment data-informed guidance to individualize kindergarten reading instruction:
Findings from a cluster-randomized control field trial. *Elementary School Journal,
111*(4), 535-560. doi:10.1086/659031

Almasi, J. F., O'Flahavan, J. F., & Arya, P. (2001). A comparative analysis of students and
teacher development in more and less proficient discussions of literature. *Reading
Research Quarterly, 36*, 96-120.

Bowers, E. P., & Vasilyeva, M. (2010). The relation between teacher input and lexical growth of
preschoolers. *Applied Psycholinguistics, 32*(1), 221-241.
https://doi.org/10.1017/S0142716410000354

Bredekamp, S. (1986). The reliability and validity of the Early Childhood Classroom
Observation Scale for accrediting early childhood programs. *Early Childhood Research
Quarterly*, *1*(2), 103-118.

Burchinal, M. R. (2018). Measuring early care and education quality. *Child Development
Perspectives, 12*(1), 3–9.

Burchinal, M. R., Peisner-Feinberg, E., Bryant, D. M., & Clifford, R. (2000). Children's social
and cognitive development and child-care quality: Testing for differential associations
related to poverty, gender, or ethnicity. *Applied Developmental Science*, *4*(3), 149-165.

CA Department of Education. (2015). *DRDP: A Developmental Continuum from Early Infancy
to Kindergarten Entry*. Retrieved from
https://www.cde.ca.gov/sp/cd/ci/documents/drdp2015preschool.pdf

Carlisle, J. F., Dwyer, J., & Learned, J. (in press). Discussion as a means of learning to reason, read, and write analytically.

Chaudry, A., & Datta, A. (2017). The current landscape for public pre-kindergarten programs. *The Current State of Scientific Knowledge on Pre-kindergarten Effects*, 5-18.

Chien, N. C., Howes, C., Burchinal, M., Pianta, R. C., Ritchie, S., Bryant, D. M., . . . Barbarin, O. A. (2010). Children's Classroom Engagement and School Readiness Gains in Prekindergarten. *Child Development, 81*(5), 1534-1549. doi:10.1111/j.1467-8624.2010.01490.x

Connor, C. M., Kelcey, B., Sparapani, N., Petscher, Y., Siegal, S. W., Adams, A., & Carlisle, J. F. (in press). Predicting Second and Third Graders' Reading Comprehension Gains: Observing Students' and Classmates Talk during Literacy Instruction using COLT. *Scientific Studies of Reading*.

Connor, C. M., Morrison, F. J., Fishman, B., Giuliani, S., Luck, M., Underwood, P., . . . Schatschneider, C. (2011a). Classroom instruction, child X instruction interactions and the impact of differentiating student instruction on third graders' reading comprehension. *Reading Research Quarterly, 46*(3), 189-221.

Connor, C. M., Morrison, F. J., Fishman, B., Ponitz, C. C., Glasney, S., Underwood, P., . . . Schatschneider, C. (2009). The ISI classroom observation system: Examining the literacy instruction provided to individual students. *Educational Researcher, 38*(2), 85-99. doi:10.3102/0013189X09332373

Connor, C. M., Morrison, F. J., Schatschneider, C., Toste, J., Lundblom, E. G., Crowe, E., & Fishman, B. (2011b). Effective classroom instruction: Implications of child characteristic

by instruction interactions on first graders' word reading achievement. *Journal of Research on Educational Effectiveness, 4*(3), 173-207. doi:10.1080/19345747.2010.510179

Connor, C. M., Morrison, F. J., & Slominski, L. (2006). Preschool instruction and children's emergent literacy growth. *Journal of Educational Psychology, 98*(4), 665-689.

Connor, C. M., Piasta, S. B., Fishman, B., Glasney, S., Schatschneider, C., Crowe, E., . . . Morrison, F. J. (2009). Individualizing student instruction precisely: Effects of child× instruction interactions on first Graders' literacy development. *Child Development, 80*(1), 77-100. doi:10.1111/j.1467-8624.2008.01247.x

Connor, C. M., Spencer, M., Day, S. L., Giuliani, S., Ingebrand, S. W., McLean, L., & Morrison, F. J. (2014). Capturing the complexity: Content, type, and amount of instruction and quality of the classroom learning environment synergistically predict third graders' vocabulary and reading comprehension outcomes. *Journal of Educational Psychology, 106*(3), 762-778. doi: 10.1037/a0035921

Day, S. L., & Connor, C. M. (2016). Examining the relations between self-regulation and achievement in third-grade students. *Assessment for Effective Intervention, 42*(2), 97-109. doi:10.1177/1534508416670367

Downer, J. T. , Booren, L. M. , Lima, O. K. , Luckner, A. E., & Pianta, R. C. (2010). The Individualized Classroom Assessment Scoring System (inCLASS): Preliminary reliability and validity of a system for observing preschoolers' competence in classroom interactions. *Early Childhood Research Quarterly*, *25*(1), 1–16.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., & Klebanov, P. (2007). School readiness and later achievement. *Developmental Psychology, 43*, 1428-1446.

Garcia Coll, C. T. (1990). Developmental outcome of minority infants: A process-oriented look into our beginnings. *Child Development*, *61*(2), 270-289.

Harms, T., Clifford, R., & Cryer, D. (2015). *Early Childhood Environment Rating Scale 3rd Edition*. New York, NY: Teachers College Press.

Hestenes, L. L., Kintner-Duffy, V., Wang, Y. C., La Paro, K., Mims, S. U., Crosby, D., ... & Cassidy, D. J. (2015). Comparisons among quality measures in child care settings: Understanding the use of multiple measures in North Carolina's QRIS and their links to social-emotional development in preschool children. *Early Childhood Research Quarterly*, *30*, 199-214.

Hill, H. C., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L., & Ball, D. L. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cognition and Instruction, 26*(4), 430-511. doi:10.1080/07370000802177235

Huttenlocher, J., Vasilyeva, M., Waterfall, H. R., Vevea, J. L, & Hedges, L. V. (2007). The varieties of speech to young children. *Developmental Psychology*, *43*(5), 1062-83. doi: 10.1037/0012-1649.43.5.1062

Lamb, M. E. (1998). Nonparental child care. In *Parenting and Child Development in Nontraditional Families* (pp. 47-64). London: Psychology Press.

Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O., Bryant, D., Burchinal, M. et. al. (2008). "Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills." *Child Development, 79*(3), 732-749.

McCartney, K., Dearing, E., Taylor, B. A., & Bub, K. L. (2007). Quality child care supports the achievement of low-income children: Direct and indirect pathways through caregiving and the home environment. *Journal of Applied Developmental Psychology*, *28*(5-6), 411-426.

McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin, 114*(2), 376.

Murphy, P. K., Wilkinson, I. A. G., Soter, A. O., Hennessey, M. N., & Alexander, J. F. (2009). Examining the effects of classroom discussion on students' comprehension of text: A meta-analysis. *Journal of Educational Psychology, 101*(3), 740-764. doi:10.1037/a0015576

National Institute of Child Health and Human Development (NICHD) Early Child Care Research Network (2000). Characteristics and quality of child care for toddlers and preschoolers, *Applied Developmental Science, 4*(3), 116-135.

Nelson, K. E., & Bonvillian, J. D. (1973). Concepts and words in the 18-month-old: Acquiring concept names under controlled conditions. *Cognition*, *2*(4), 435-450.

Pianta, R. C. & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher*, *38*(2), 109–119. doi: 10.3102/0013189X09332374

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

Ritchie, S., Howes, C., Kraft-Sayre, M., & Weiser, B. (2001). *Emerging Academic Snapshot*. Unpublished measure, University of California at Los Angeles.

Smolak, L., & Weinraub, M. (1983). Maternal speech: Strategy or response? *Journal of Child Language, 10*(2), 369-380.

Snow, C. E. (1989). Understanding social interaction and language acquisition: Sentences are not enough. In M. Borstein & J. Bruner (Eds.), *Interaction in human development* (pp. 83-104). Hillsdale, NJ: Lawrence Erlbaum.

Stipek, D., & Byler, P. (2004). The Early Childhood Classroom Observation Measure. *Early Childhood Research Quarterly*, *19,* 375–397.

Teaching Strategies. (2018). The Creative Curriculum for Preschool. Retrieved from https://teachingstrategies.com/solutions/teach/preschool/

VORT. (2018). *Hawaiian Early Learning Profile*. Retrieved from https://www.vort.com/

Table 1

Demographic Information by Program

	State PK	State PK – Special Education	Head Start	Tuition-based PS
Target population	Higher poverty and other risks	Higher poverty or special needs without regard to income	Higher poverty	Open to all; located on a university campus
Children (n)	23 assessed, 18 observed	38 assessed (about 80% IEPs), 14 observed	28 assessed, 27 observed	19 assessed, 9 observed
Teachers/Classrooms (n)	4 half day	3 half day	4 half day	1 full day
Teaching assistants or other adults in each classroom (n)	1	2-3	1	4-6
Average amount of time observed in minutes	188 (SD=17,167-212)	196 (SD=113, 125-327)	237 (SD=89,101-315)	288 (SD=N/A)
Mean age (years; months)	4;8	4;10	4;6	5;3
Funding	State	State and Special education	Federal	Tuition, non-profit
Vocabulary (DSS)*	451	457	456	475
Early Literacy (GE)*	K.0	K.2	K.2	K.7
Other languages?	Spanish, Arabic, Vietnamese, and others	Spanish	Spanish and Vietnamese	None

* Spring assessment.

Note. DSS= Developmental Scale Score. GE=Grade Equivalent. Vocabulary is a measure of vocabulary depth; Early Literacy is a measure of letter and letter sound knowledge, and word reading. A GE of -1.0 would indicate a child at the beginning of preschool, K.0 would indicate a child at the beginning of kindergarten, 1.0 would indicate a child at the beginning of first grade, etc.

Table 2

Descriptive statistics for the amount of time in different learning activities (minutes observed per 30 minutes) on top, and mean frequency of child talk factors (frequency observed per 30 minutes) on the bottom

	Mean	Minimum	Maximum	Std. Deviation
Time in activity				
Literacy	9.18	.00	21.64	5.46
Math	.34	.00	9.33	1.33
Science	.83	.00	12.66	2.63
Arts	3.84	.00	20.16	4.89
Play	4.88	.00	14.99	5.24
Off task	1.83	.00	30.00	6.10
Non-instruction and meals	9.12	.00	28.62	7.77
Frequency				
Child Talk	6.34	.00	43.37	9.92
Social Behavior	.52	.00	21.69	2.67
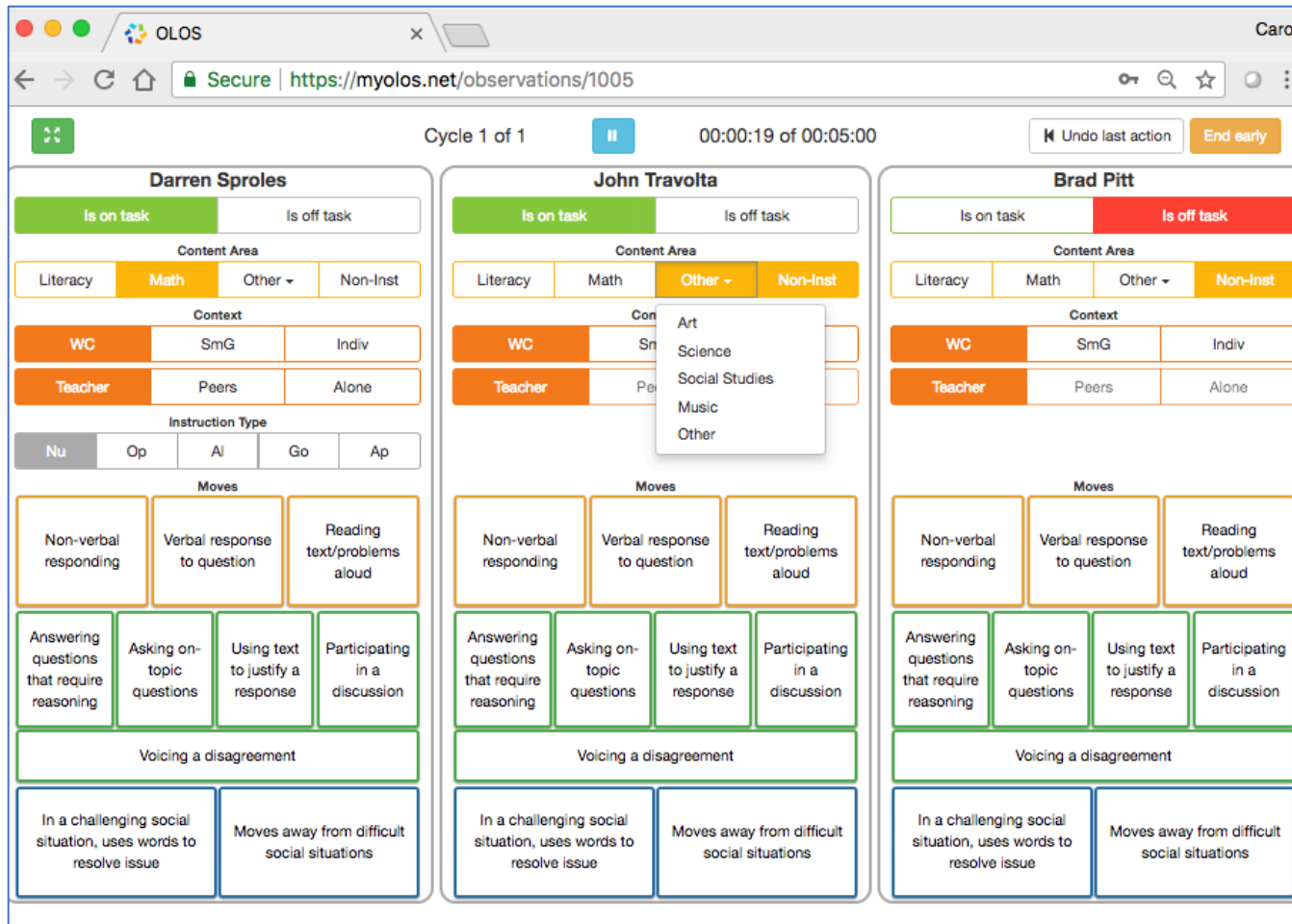
Figure 1. OLOS Child Coding Interface screenshot. Names are pseudonyms. The coder has coded Darren as on task, participating in a whole class (WC) activity with the teacher focused on numbers (NU). John is on task and the reviewer is about to select what the content of instruction is. Brad is off task and is not involved in instruction (non-instruction). Other contexts including small group

(SmG) and individual (Indv). Students can be with the teacher, with peers, or working alone. Time is recorded until the activity changes. For the discourse codes, each time the button is pressed, one event is recorded (e.g., non-verbal responding). The coding manual is available upon request from the corresponding author.

Figure 2. OLOS Teacher Coding Interface screenshot. The coding manual is available upon request from the corresponding author.

Cycle A1	Child A,B,C 15 min	Teacher 5 min	Child A,B,C 15 min	Teacher 5 min
Cycle A2	Child A,B,C 15 min	Teacher 5 min	Child A,B,C 15 min	Teacher 5 min
Cycle B1	Child D,E,F 15 min	Teacher 5 min	Child D,E,F 15 min	Teacher 5 min
Cycle B2	Child D,E,F 15 min	Teacher 5 min	Child D,E,F 15 min	Teacher 5 min
Cycle C1: etc.				

Figure 3. Procedures for observing classrooms using OLOS through repeating cycles.
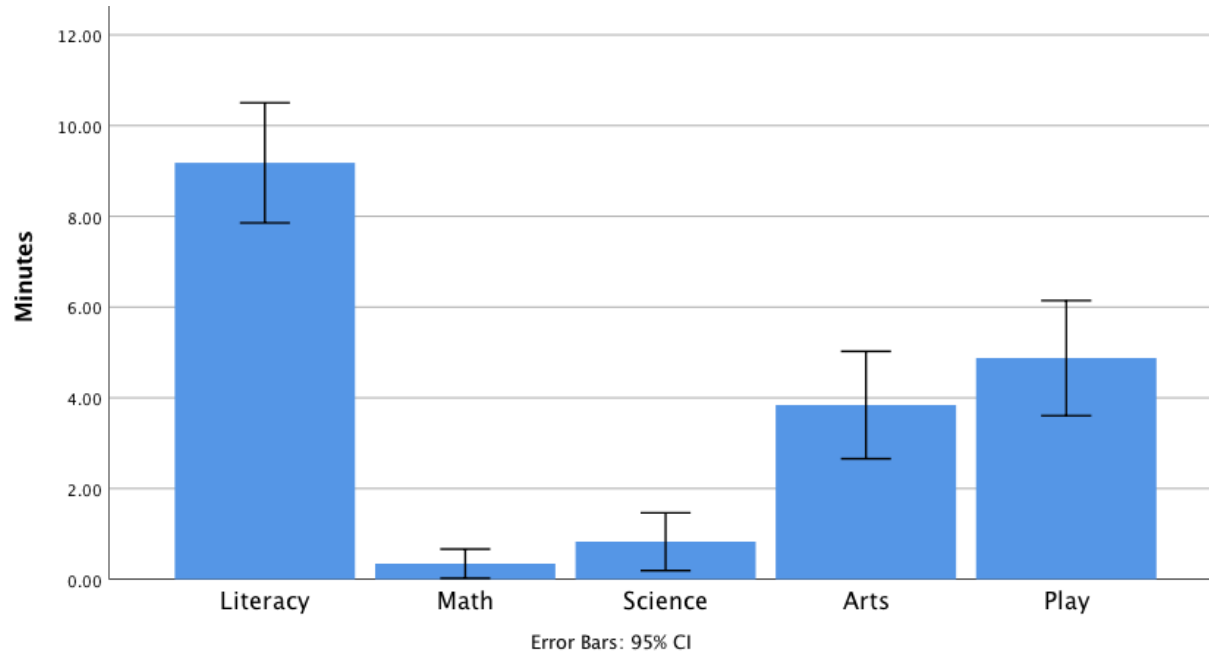
Figure 4. Mean time (min) children spent in various types of learning opportunities during 30 minutes of observation. Error bars are

95% confidence intervals.
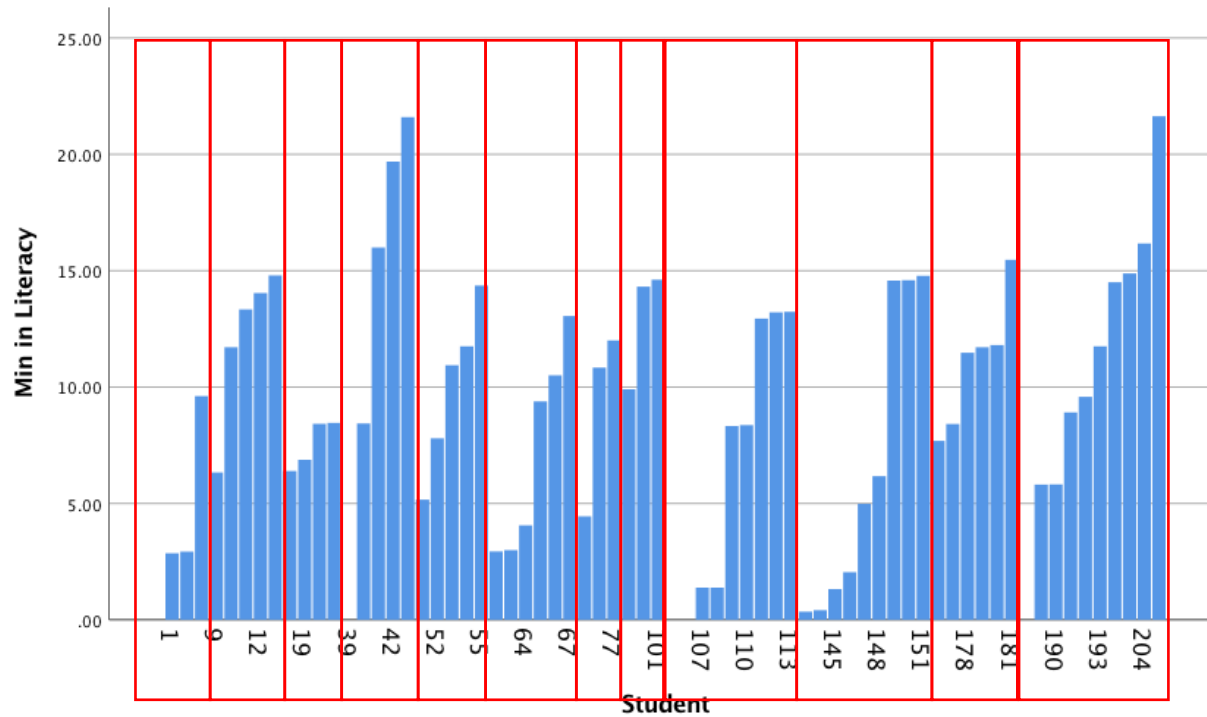
Figure 5. Each bar represents the number of minutes/30 minutes each child spent in literacy learning opportunities. Students sorted by

classroom, program, and time spent in literacy learning opportunities. Within classroom variability can be seen by comparing

contiguous bars within each box (e.g., student 1-8 are in the same classroom).
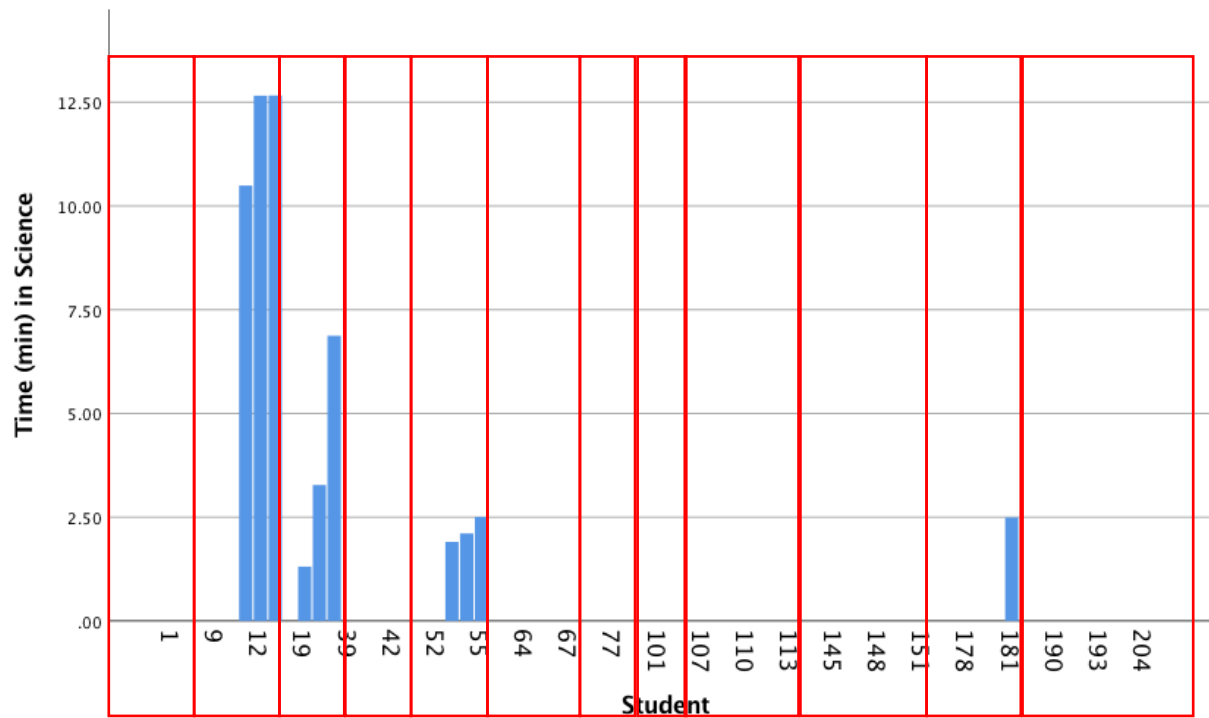
Figure 6. Each bar represents one child's time spent in science learning activities (min/30 min). Data are sorted by classroom, program type, and time in science learning activities. Each box represents one classroom.
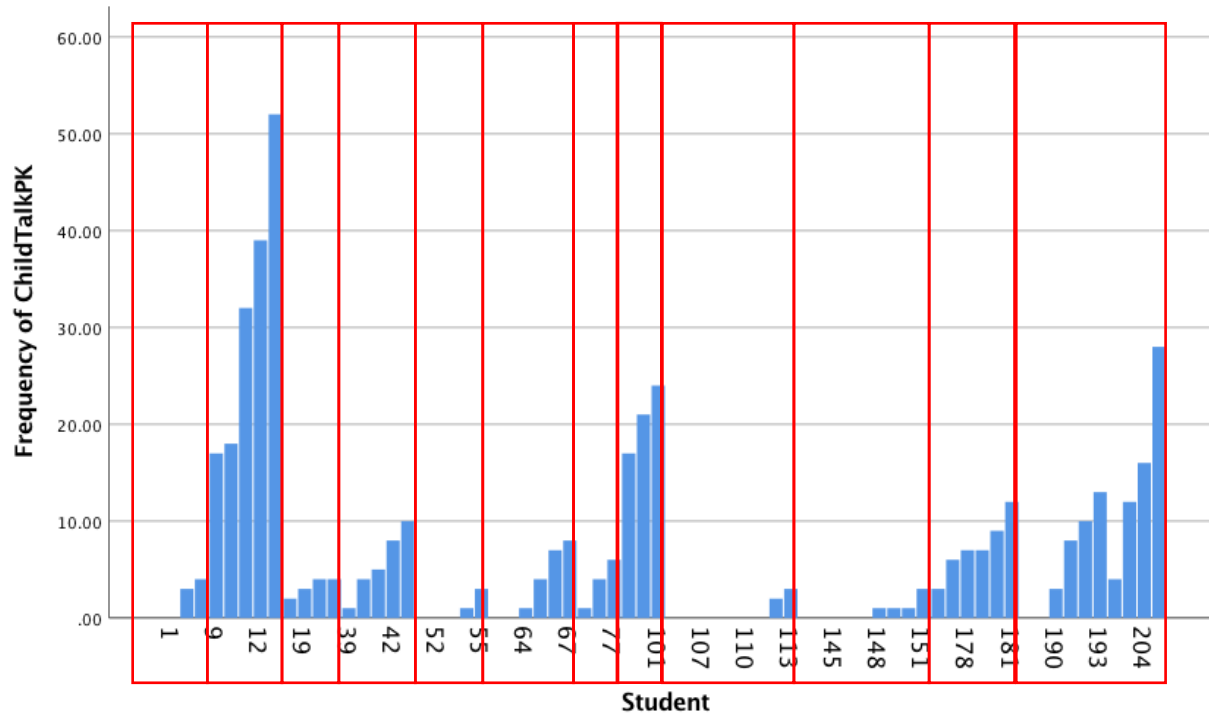
Figure 7. Each bar represents the frequency of Child Talk per 30 minutes for an individual child, sorted by classroom, program, and

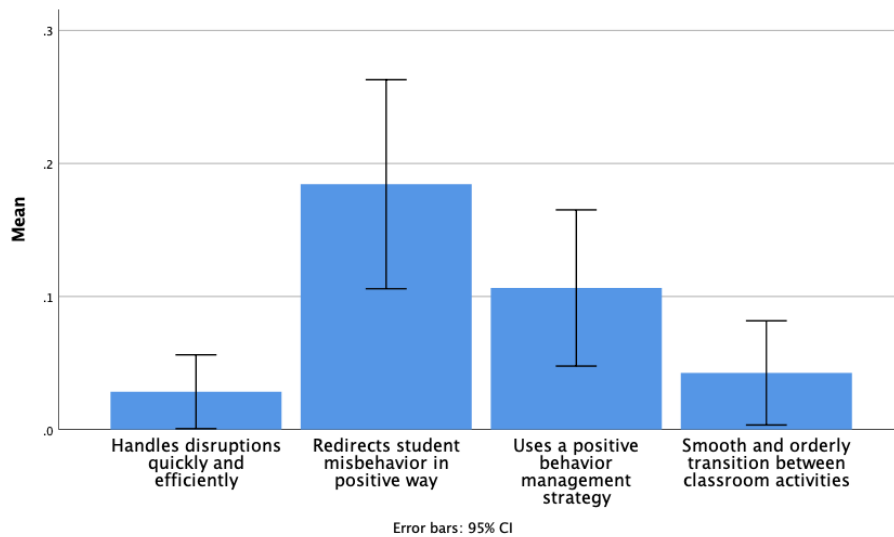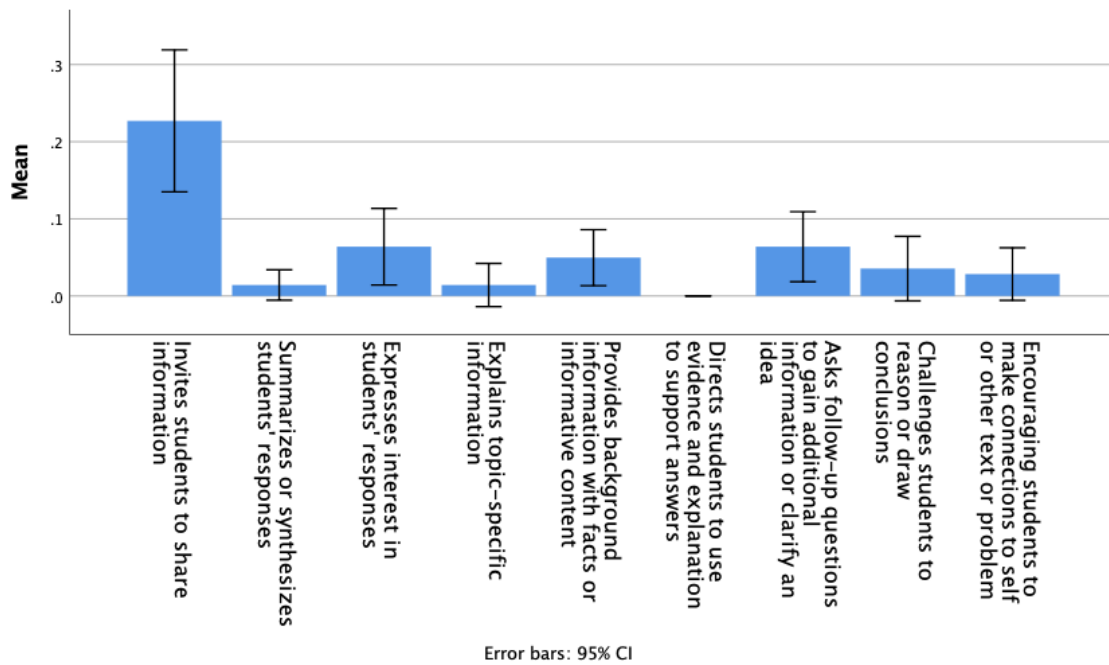frequency of Child Talk. Each box represents one classroom.

Figure 8. Mean frequencies of teacher talk types (top) and social-behavioral support actions (bottom).
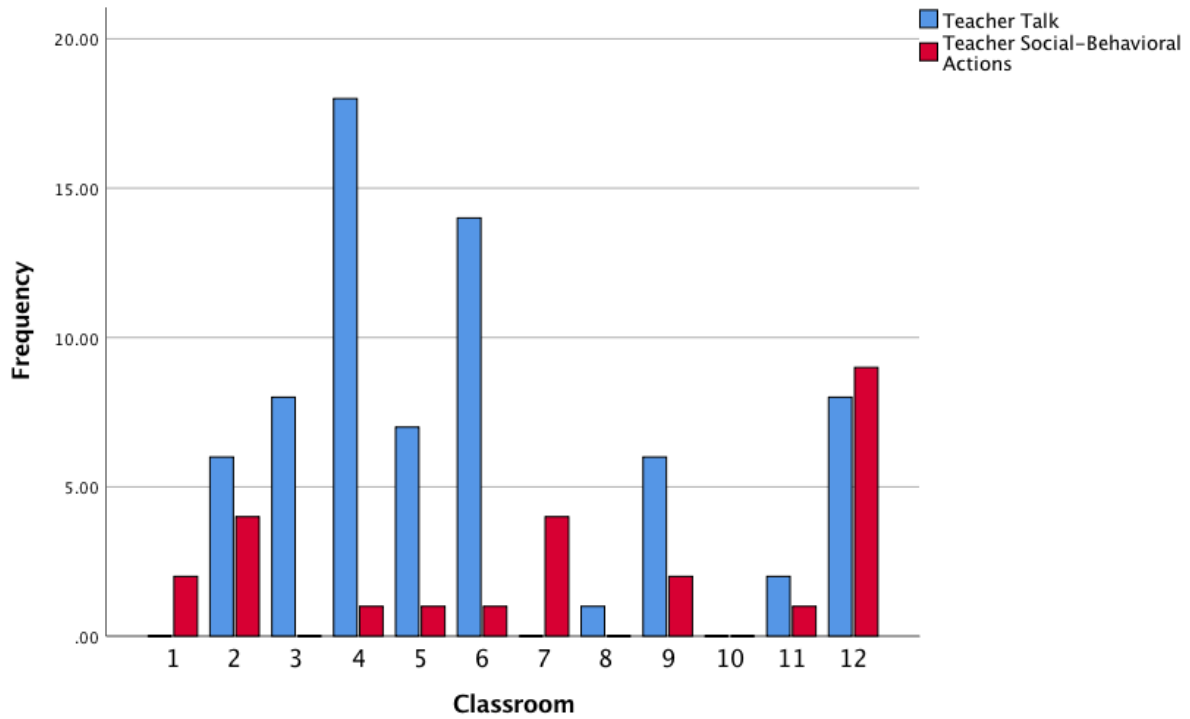
Figure 9. Total frequency of Teacher Talk and Social-Behavioral Actions observed for each

teacher.

Figure 10. Random selection of students by grade level restriction functionality.

Appendix

Table A.1

Factor Matrix for Child Talk Frequency

	1	2	3	4
Answering questions that require reasoning	.757	.017	-.653	.000
Verbal response to question	.653	.244	.061	-.081
Participating in a discussion	.369	-.125	-.302	-.068
Asking on-topic questions	.546	-.739	.394	.000
Non-verbal responding	.600	.650	.465	.000
Voicing a disagreement	.415	-.483	.281	-.017
Reading texts/problems aloud	.267	.425	.250	-.139
Used words to resolve social issues	-.105	.038	.038	.648
Moves away from difficult social situations	-.014	.043	-.058	.587

Extraction Method: Generalized Least Squares.
a. 4 factors extracted. 23 iterations required.