Early Childhood Research Quarterly 24 (2009) 213–228

Contents lists available at ScienceDirect

Early Childhood Research Quarterly

Inside the pre-kindergarten door: Classroom climate and instructional time allocation in Tulsa's pre-K programs

Deborah A. Phillips*, William T. Gormley, Amy E. Lowenstein

Georgetown University, United States

ARTICLE INFO

Article history:
Received 1 March 2008
Received in revised form 8 May 2009
Accepted 14 May 2009

Keywords:
Pre-kindergarten
Early childhood education
Quality
Classroom
Climate
Instructional time

ABSTRACT

This observational study of 106 pre-K classrooms in Tulsa, Oklahoma provides descriptive data on children's classroom experiences – classroom climate and exposure to academic instruction – and comparisons of Tulsa classrooms with a multi-state sample of pre-K and Head Start classrooms led by comparably educated teachers. We also examined teacher and classroom characteristics that were associated with variation in children's classroom experiences. Compared to other school-based pre-K and Head Start classrooms, teachers in the Tulsa classrooms received significantly higher scores for various dimensions of Instructional Support and Classroom Organization, and devoted significantly more time to academic instruction, notably Literacy and Math Activities. Within the Tulsa sample, children in Head Start classrooms received less exposure to Math instruction, but more exposure to Social Studies activities as compared to their peers in public school pre-K classrooms. Teacher and classroom characteristics were not associated with classroom climate, and only the teachers' years of experience was associated with greater exposure to Literacy Activities. Major challenges remain in identifying classroom and teacher attributes that contribute to high-quality, educationally rich pre-K environments.

1. Introduction

The United States is fast approaching a time when school will start at age four for the majority of children. State-funded pre-K programs are now offered in 38 states and the District of Columbia, serving 28% of the nation's 4-year-olds (Barnett, Epstein, Friedman, Boyd, & Hustedt, 2008). This rapid growth in state pre-K programs represents a national experiment focused on finding the best means of launching all young children on a trajectory of school success. Developmental scientists are interested in the impacts of this phenomenon on children's short-term academic and social development, as well as on their growth curves. Policy makers share this primary interest in impacts, but also face practical decisions about how to focus public resources on those design, organizational, and staffing attributes of pre-K programs that will produce high-quality classroom experiences for children and thus hold the most promise of generating strong impacts. This study addresses this practical question in the context of Oklahoma's pre-K program, which has generated strong impacts across racial and income groups in three cohorts of students (Gormley & Gayer, 2005; Gormley, Gayer, Phillips, & Dawson, 2005; Gormley, Phillips, & Gayer, 2008).

* Corresponding author at: Department of Psychology, Georgetown University, White Gravenor, rm. 301F, 3700 O St., NW Washington, DC 20057, United States. Tel.: +1 202 687 4042.
E-mail address: Deborah.dap4@gmail.com (D.A. Phillips).

0885-2006/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.ecresq.2009.05.002

1.1. Variation in state-funded pre-K programs

Emerging descriptive data indicate that state-funded pre-K programs are characterized by extensive variation. This variation starts with state pre-K standards. For example, while 26 states require the BA degree and early childhood certification for all pre-K lead teachers, eight states do not require any teachers to have a BA degree and 10 states do not require teachers to have specialized training in early childhood education. The majority of states require a teacher–child ratio of 1:10, but 12 states allow less stringent ratios in some or all of their pre-K classrooms (Barnett et al., 2008).

Oklahoma has among the most stringent state pre-K standards, with its requirement that every classroom's lead teacher have a BA degree and an early childhood teaching certificate, and a maximum class size of 20 and teacher–student ratio of 1:10. These standards apply not only to pre-K programs based in the Tulsa Public Schools (TPS), but also to collaborating Head Start and childcare programs. Oklahoma also pays its pre-K teachers the same wages that it pays other public school teachers. The Tulsa Head Start program has chosen to pay its teachers similar wages in order to compete with the TPS schools for top teachers. While 13 programs, like Oklahoma, require that all pre-K teachers be paid on the public school salary scale, many add caveats to this requirement (e.g., it applies only to certified or BA-level teachers) (Barnett et al., 2008).

Available, albeit sparse, evidence suggests that variation in pre-K programs extends beyond state standards to widely disparate experiences for children within these settings (Clifford et al., 2005; Early et al., 2006; Gilliam & Marchesseault, 2005; LoCasale-Crouch et al., 2007). In two recent reports of pre-K practices in 11 states (Early et al., 2006; Howes et al., 2008), average classroom quality assessed with the Early Childhood Environment Rating Scale-Revised (ECERS-R) hovered just below the "good" level, with 12% of classrooms receiving quality scores in the minimal range and 8% receiving scores in the good to excellent range (Early et al., 2006). Moreover, while the quality of emotional interactions between teachers and children in these classrooms fell within the middle-to-high range on the Classroom Assessment Scoring System (CLASS) (Pianta, La Paro, & Hamre, 2008), the vast majority of classrooms fell within the low-quality range on the quality of instructional interactions (Early et al., 2006; Howes et al., 2008). Children spent substantially less than 10% of their classroom time on specific pre-literacy and pre-math activities.

A representative sample of Head Start classrooms has also been assessed with the ECERS-R (U.S. Department of Health and Human Services, 2006). Average classroom quality was notably higher than that reported in the 11-state pre-K sample; the vast majority of classrooms fell within the "good" range. Nevertheless, the proportion of time spent on early literacy activities was minimal, as with the pre-K classrooms. Importantly, these data were collected in 2003, prior to the Congressional mandate to increase the share of more highly educated teachers in Head Start classrooms. In the current study, we examine whether Oklahoma's pre-K program, with its relatively stringent state requirements, is characterized by higher quality classroom environments and experiences as compared to larger and more typical pre-K and Head Start samples.

1.2. Children's experiences in pre-K classrooms

A second purpose of this study is to examine the features of the Tulsa pre-K programs that are associated with higher quality environments and experiences *within* Tulsa classrooms. Evidence is only now emerging from efforts to look inside the "black box" of pre-kindergarten classrooms to identify the specific practices and characteristics of programs that contribute to the promising results that the Oklahoma program and others have been reporting (see Barnett et al., 2008; Gilliam & Zigler, 2001, 2004; Henry, Gordon, Mashburn, & Ponder, 2001; Magnuson, Meyers, Ruhm, & Waldfogel, 2004; Pianta et al., 2005; Wong, Cook, Barnett, & Jung, 2008). Pianta and colleagues have reported positive associations between the emotional and instructional climate of the classroom and children's early learning and social-emotional development (Hamre & Pianta, 2005; Mashburn et al., 2008). More generally, the pre-K literature has directed attention to the importance of the teacher's capacity to combine focused time spent on explicit subject-matter learning, clear and efficient time management and classroom organization, and a classroom climate characterized by warm, contingent interactions (Bowman, Donovan, & Burns, 2000; Clements, Sarama, & DiBiase, 2003; Connor, Morrison, & Slominski, 2006; Howes & Richie, 2002; Snow, Burns, & Griffin, 1998; Stipek & Byler, 2003). Howes et al. (2008), for example, found that children's exposure to pre-literacy activities, the overall instructional climate of the classroom, and teacher-reported close relationships with the children were associated with higher scores on tests of early literacy at the conclusion of the pre-K year.

These findings are similar to those emerging from the Head Start Family and Child Experiences Survey (FACES) data showing that, while higher quality classrooms are led by teachers with higher levels of education (especially a BA or AA degree) and a teaching certificate, these characteristics affect classroom quality through their positive influence on teacher attitudes and knowledge (Resnick & Zill, 2003). In effect, this relatively new line of research is confirming long-standing evidence on the significant role played by *process* dimensions of classroom quality in supporting early learning.

These findings from relatively recent studies of pre-K and Head Start programs are also consistent with a longer standing literature examining predictors of high-quality instruction and student performance in elementary school. The elementary school research is restricted to samples of teachers with BA degrees, given consistent state standards on teacher qualifications and, as such, is highly relevant to the pre-K context in Oklahoma (where every lead teacher must have a BA degree). In this literature, neither higher levels of education in the form of master's degrees nor certification per se are associated with stronger student performance, in part because these variables do not appear to distinguish good teachers from bad (Aaronson, Barrow, & Sander, 2003; Boyd, Goldhaber, Lankford, & Wyckoff, 2007; Clotfelter, Ladd, & Vigdor, 2006; Ehrenberg & Brewer, 1997; Hill, 2007; Kane, Rockoff, & Staiger, 2006). Positive associations with elementary student performance are,

however, found for teacher experience and ability. More experienced teachers provide higher quality classroom instruction and, in turn, produce stronger student achievement in both math and reading (Clotfelter et al., 2006; Goldhaber, 2007; Jacob, 2007; Kane et al., 2006; Rivkin, Hanushek, & Kain, 2005; Rockoff, 2004). This link between teacher experience and positive classroom experiences has also been found in the pre-K literature (LoCasale-Crouch et al., 2007; Pianta et al., 2005). Some of this experience effect appears to be associated with the relatively poorer skills of novice teachers who are in their first or second year of teaching.

Teacher ability has been assessed using the selectivity of the college a teacher attended (Summers & Wolfe, 1977) and performance on various tests ranging from certification and college entrance exams to tests of verbal ability (Clotfelter et al., 2006; Ehrenberg & Brewer, 1997; Ferguson, 1991; Hanushek, 1971). While these proxies for teachers' cognitive ability are more strongly linked to student performance than are degrees and credentials, the associations remain quite weak (Hanushek & Rivkin, 2004).

An additional line of research aimed at deciphering effective early educational and elementary practice has focused on curricula. Despite repeated attempts to distinguish curricula that most strongly support early learning, results have been inconsistent and often disappointing (Cole, Mills, Jenkins, & Dale, 2005; Ehri, Nunes, Stahl, & Willows, 2001; Preschool Curriculum Evaluation Research Consortium, 2008; Ryder, Burton, & Silberg, 2006; Schweinhart & Weikart, 1997; Stebbins, St. Pierre, Proper, Anderson, & Cerva, 1977). The salient, cautionary message from this literature is that it is not about the curriculum per se, but rather about how effectively the teacher deploys the curriculum in the context of student–teacher instructional interactions and the overall emotional climate of the classroom (Phillips & Stipek, 1993; Ryder et al., 2006). In other words, it is not about the script, it is about the actors.

Taken together, the relatively new evidence on early childhood classrooms and the longer standing literature on elementary education have directed attention to a somewhat new mix of classroom features that warrant examination in future research: (1) teacher certification, (2) teacher's years of higher education, (3) teacher's years of classroom experience, (4) indicators of teacher cognitive ability, and (5) curricula used in the classroom. These features, in turn, appear to affect children's learning through their associations with the emotional and instructional climate in the classroom, and the amount of classroom time that is devoted to specific instructional activities.

1.3. The current study

The current study offers a look inside the classroom door of pre-K programs in the relatively stringent policy context of Tulsa, Oklahoma. We examined whether this policy context is associated with relatively positive classroom climates and more time on academic instruction for young children both within Tulsa and in comparison to programs in other states. Moreover, we were able to examine the role of auspice – TPS vs. Head Start – as it is associated with children's classroom experiences. This has become an increasingly pressing question in light of proposals that range from better aligning Head Start practices with those of state pre-K programs to diverting funding for 4-year-old programs from Head Start to state pre-K (Currie, 2001; Haskins & Sawhill, 2003; Ripple, Gilliam, Chanana, & Zigler, 1999).

We address three questions: (1) What are children's experiences with regard to classroom climate and exposure to academic instruction in Tulsa's pre-K classrooms? (2) Are the Tulsa pre-K classrooms characterized by more positive classroom climates and greater instructional time for 4-year-old children as compared to multi-state assessments of pre-K and Head Start classrooms? and (3) are program auspice (Head Start or TPS), classroom characteristics, and teacher characteristics and practices associated with variation in classroom processes within Tulsa? We examine virtually all the variables identified above, namely: (1) whether the teacher majored in early education (rather than certification in early childhood education, which all teachers in Tulsa are required to hold), (2) whether the teacher has a master's degree, (3) the teacher's years of classroom experience, (4) the teacher's undergraduate Grade Point Average (GPA) (a proxy for ability level), and (5) curricula used in the classroom. We examine their association with the teacher's ability to establish a positive classroom climate and the time s/he spends on instructional activities. We also examine whether there are differences in classroom climate or instructional time for half- and full-day classrooms.

2. Method

The current project is designed to examine classroom processes that may offer insights into the successful outcomes of the Tulsa pre-K program. Classroom climate and exposure to academic instruction are the processes of interest given evidence of their association with young children's learning. We used observational measures that have been employed in prior research to permit comparisons between Tulsa classrooms and other samples of pre-K and Head Start classrooms. We also gathered data from teachers and school records to explore correlates of classroom climate and exposure to academic instruction. We first describe the methods used for the Tulsa data collection, followed by a description of the multi-state comparison datasets.

2.1. Participating classrooms and teachers

The focus of our inquiry is pre-K classrooms run by the Tulsa Public Schools (TPS) and 4-year-old Head Start classrooms run by the Community Action Project (CAP) of Tulsa County. Both sets of classrooms are funded by the state of Oklahoma,

under a universal pre-K system established in 1998. The Head Start classrooms are eligible for funding because they have established a "collaborative" relationship with Tulsa Public Schools and agreed to meet the pre-K quality requirements of the TPS school system. As of fall 2006, all 4-year-old Head Start programs run by the CAP of Tulsa County were collaboratives.

At the outset of our investigation, we identified 100 pre-K classrooms and 29 Head Start classrooms that might be suitable for classroom visits. Because pre-K classrooms might be expected to differ from morning to afternoon, we decided to focus exclusively on morning experiences. For this reason, we excluded afternoon-only classrooms and afternoon sessions of full-day programs from consideration, thus leaving a total of 80 pre-K classrooms and 29 Head Start classrooms eligible for observation. We did, however, code each classroom according to whether it was half- or full-day. We also excluded blended classrooms (i.e., classrooms that included both 3- and 4-year-old) from our Head Start sample, as these classrooms might be expected to differ from those that serve 4-year-old exclusively. Of 59 full-day TPS classrooms, our student observers visited 58, always in the morning. Of 21 morning TPS classrooms, our observers visited 20. Of 29 Head Start classrooms, our observers visited 28, always in the morning. Missing observations were due to scheduling problems. In short, our sample constitutes virtually the entire universe of state-funded morning classrooms for 4-year-old in Tulsa, Oklahoma and 106 of the 129 total 4-year-old classrooms in Tulsa.

As with all pre-K programs in the state of Oklahoma, the programs in Tulsa were required to employ a lead teacher with a BA degree and an early childhood certification, which encompasses instruction for children in pre-K through third grade, and to maintain a 10:1 student–teacher ratio. Maximum group size across the state is 20 children and, as a result, classrooms typically include one assistant teacher for whom there are no educational or training requirements. Across the state as a whole, 97% of lead teachers had a BA degree or better and 97% had their early childhood certification, and 20% of classroom ratios exceeded the state mandate of 10:1 (Gilliam & Marchesseault, 2005). In Tulsa, 100% of lead teachers had a BA degree or better and 94% (96% of TPS teachers and 89% of Head Start teachers) had their early childhood certification. We did not collect data on child–teacher ratios, although our observers did report informally that the number of children present exceeded the group size maximum of 20 children on rare occasions.

Of the 106 lead teachers in the 106 participating classrooms, 104 completed questionnaires about their educational background and training, employment history, and use of pre-K curricula. Seventy-seven of the teachers who completed questionnaires provided consent for us to obtain a copy of their college transcript. These 77 teachers differed somewhat from the 27 teachers who did not give us permission to obtain their transcript on a number of variables of interest. Specifically, there were marginally significant differences in the teachers' educational backgrounds, such that teachers who provided us with their transcript were more likely to have a BA in early childhood education (vs. some other field) and less likely to have an MA (both p's < .10). Teachers who provided us with their transcript were also more likely to teach in a half-day classroom ($p < .01$). The two groups of teachers did not differ in their years of teaching experience or in the proportion that were Head Start teachers. The vast majority of teachers in our sample were female (102), 28 were in their first or second year of teaching, and most had had prior experience teaching in a pre-K or Head Start classroom.

2.2. Procedure and measures

Between February and May of 2006, classroom observations were conducted by pairs of trained observers in the 106 participating classrooms. Observers were undergraduate students at the University of Tulsa who had received classroom instruction in child development and/or early education. Each pair spent the morning, from student arrival until dismissal for lunch, observing each classroom. Observations were conducted continuously. Lead teachers were given a questionnaire during the classroom visits to be completed on site or returned post-visit.

2.2.1. Classroom observations

Two instruments were used to observe the pre-K classrooms: the *Classroom Assessment Scoring System* or CLASS (La Paro, Pianta, Hamre, & Stuhlman, 2002; La Paro, Pianta, & Stuhlman, 2004; Pianta et al., 2008) was used to capture the instructional and emotional climate in the classrooms, as well as the teachers' skill at classroom organization and their students' degree of engagement in classroom activities; and the *Child Engagement section of the Emerging Academics Snapshot* or CE-EAS (Ritchie, Howes, Kraft-Sayre, & Weiser, 2002) was used to capture exposure to academic instruction. The former is generally portrayed as a measure of classroom quality, whereas the latter is most accurately viewed as a measure of the quantity of academic instruction that may or may not be conveyed in a high-quality fashion. We consider both to capture classroom processes and thus children's experiences, with the CLASS focusing on *how well* the classroom is functioning and the CE-EAS focusing on *how much* exposure to academic material is provided. Due to illness, we were unable to administer the CLASS in one TPS classroom and the CE-EAS in another TPS classroom. Thus, our working sample for both the CLASS and the CE-EAS is 77 TPS classrooms and 28 Head Start classrooms.

The CLASS provides an assessment of 11 dimensions of classroom climate (see below). Each dimension is rated on a scale of 1–7, with 1–2 indicating low levels, 3–5 medium levels, and 6–7 high levels. Each cycle of observation consisted of a 20-min period of watching classroom interaction and taking notes, followed by a 10-min recording period. In the majority of classrooms, the CLASS observer completed five of these cycles, excluding only recess time. A score for each dimension was created by averaging ratings across all cycles. Additionally, using the CLASS dimensions, we created three composites based on the results of a factor analysis conducted by scholars using data from the Multi-State and SWEEP states, as well as other sources (see Hamre, Mashburn, Pianta, LoCasale-Crouch, & La Paro, n.d.). These three factors are: *Emotional Support*

Table 1
Means[a] and standard deviations for CLASS, CE-EAS, classroom, and teacher variables by TPS pre-K/CAP Head Start status.

Variable	TPS pre-K ($n = 77$)[b]		CAP Head Start ($n = 28$)[c]	
	M	S.D.	*M*	S.D.
Emotional support composite (CLASS)	5.23	0.57	5.22	0.78
Positive climate	5.06	0.79	4.98	0.92
Negative climate	1.36	0.56	1.50	0.74
Teacher sensitivity	4.83	0.67	4.76	0.94
Regard for student perspectives	4.37	0.81	4.62	0.98
Classroom organization composite (CLASS)	4.96	0.69	4.80	0.84
Behavior management	5.03	0.85	4.45	1.28
Productivity	5.21	0.84	5.05	0.98
Instructional learning formats	4.64	0.87	4.90	0.74
Instructional support composite (CLASS)	3.21	0.93	3.26	0.94
Concept development	2.83	1.19	2.59	1.23
Quality of feedback	3.29	0.98	3.48	1.03
Language modeling	3.51	0.98	3.72	1.06
Student engagement (CLASS)	5.23	0.71	5.09	0.77
Exposure to academic instruction dimensions (CE-EAS)				
Literacy activities composite	.30	.13	.24	.10
Writing	.03	.03	.02	.03
Engaged in math	.17	.10	.11	.05
Engaged in science	.17	.10	.18	.10
Engaged in social studies	.13	.08	.23	.08
Engaged in aesthetics	.19	.10	.23	.14
Classroom characteristics/curricula				
Half-day classroom	.23	.42	.00	.00
Teacher uses direct instruction curriculum	.30	.46	.14	.36
Teacher uses integrated thematic instruction curriculum	.57	.50	.18	.39
Teacher uses creative curriculum	.04	.20	.39	.50
Teacher characteristics				
Teacher has BA in early childhood education	.57	.50	.57	.50
Teacher has MA	.16	.37	.21	.42
Total decades as a classroom teacher	0.89	0.74	0.74	0.87
Undergraduate GPA	3.15	0.39	3.22	0.32

[a] CLASS means are scale scores ranging from 1-7; CE-EAS means are proportions of time spent on each activity.
[b] For most variables, the TPS $n = 77$. Exceptions are: half-day classroom ($n = 78$), curriculum and teacher variables other than teacher has MA ($n = 76$), teacher has MA ($n = 75$), and undergraduate GPA ($n = 57$).
[c] For all but one variable, the CAP Head Start $n = 28$. The exception is: undergraduate GPA ($n = 20$).

(a composite of Positive Climate, Negative Climate [reversed], Teacher Sensitivity, and Regard for Student Perspectives; alpha = .83), *Classroom Organization* (a composite of Behavior Management, Productivity, and Instructional Learning Formats; alpha = .72), and *Instructional Support* (a composite of Concept Development, Quality of Feedback, and Language Modeling; alpha = .83). One dimension, *Student Engagement*, which captures the extent of student participation in classroom activities, is not part of a composite score and is thus examined separately.

The CE-EAS provides detailed information on the time children spend engaged in specific academic activities, such as reading and pre-reading activities, expressive language development, math, science, and social studies. The second observer in each classroom focused on four children, selected at random, for a period of 20 min. She devoted 1 min to each child, observing for 20 s and coding for 40 s, and then turned to the next child. Once the 20-min cycle was completed, the observer shifted to another group of four children, selected at random, and the process began again. The observer coded throughout all activities, including recess, unless the observed child left the classroom or playground. In the majority of classrooms, the CE-EAS observer completed six of these 20-min cycles.

The Child Engagement (CE) section includes 14 individual activity codes, of which we focus on 5 (see Table 1), selected to capture individual children's exposure to instruction and activities in a range of skill domains that are rated as present or absent within the 20-s observation period (see Ritchie et al., 2002). For example, *Social Studies* captures structured activities and fantasy play that involve learning about "the children's world," including learning about cooperation and sharing, cultural diversity and family practices, and religion. *Aesthetics* is defined as art and music activities. In addition to the five individual activity codes, we also used a composite score for *Literacy Activities*, which was developed as part of the NCEDL Multi-State and SWEEP studies (D. Early, personal communication, 11 June 2008). This composite score reflects the proportion of time spent engaged in one or more of the following activities: *Being Read To, Reading, Practicing Letters and Sounds*, and *Building Expressive Language*. Because the total observational time varied across classrooms, and to facilitate comparisons with other samples, we converted each score to proportions or percentages of the observation period. Importantly, these categories are not mutually exclusive, so an estimate of total instructional time cannot be determined by summing the proportions.

2.2.2. Observer training and reliability

In January 2006, we trained a team of eight University of Tulsa students on the CLASS and CE-EAS observational measures. Trainers were experts who have been closely involved in the development and use of the CLASS and EAS measures and who trained the observational teams from the NCEDL Multi-State and SWEEP studies. Prior to data collection, reliability was established by comparing our observer ratings with expert ratings ("gold standards") on the two measures through videotaped observations. Mid-point inter-observer reliability assessments were conducted to ensure that these reliability levels were maintained. For the CLASS, the mid-point assessment consisted of a mix of live, double-coded observations among the observers and coding of expert-rated videotapes. For the CE-EAS, this assessment consisted exclusively of coding expert-rated videos. Because the CLASS is coded on a 7-point ordinal scale, inter-rater reliability was established at both time points through the calculation of intraclass correlation coefficients. Observers' mean intraclass correlation coefficients were 0.78 prior to data collection and 0.69 at mid-point, demonstrating adequate to good levels of inter-rater agreement. For the CE-EAS, inter-rater reliability was established prior to data collection through the calculation of percent agreement, and at mid-point through the calculation of weighted Kappas. Prior to data collection, on average, 91% of observer responses were exactly the same as the expert's responses, and all observers achieved at least 89% reliability. At mid-point, observers' mean weighted Kappa was 0.85, indicating high levels of agreement.

2.2.3. Teacher questionnaires and transcripts

Each lead teacher in the participating TPS and Head Start classrooms was provided with a questionnaire that inquired about educational background and training, teaching experience (in general, and specific to pre-K and/or Head Start), and use of specific pre-K curricula. For the present study, we created variables for whether the teacher had a BA in early childhood education (vs. a BA in education, human development, or some other field), whether or not the teacher had an MA, total years (in decades) of classroom instructional experience, and, based on the most commonly used curricula, use of the Direct Instruction, Integrated Thematic Instruction, and Creative Curriculum curricula. At the end of the questionnaire, teachers were asked to provide consent for us to obtain their undergraduate transcripts, from which we obtained information about their GPA.

2.3. Comparison classrooms

In order to place the descriptive data from the Tulsa classrooms in a broader context, the National Center for Early Development and Learning (NCEDL) Multi-State Study of Pre-Kindergarten and State-Wide Early Education Programs (SWEEP) Study (Early et al., 2005) provided comparison classroom data for 11 states. Data collection for the Multi-State Study took place during the 2001–2002 school years in six states: California, Georgia, Illinois, Kentucky, New York, and Ohio. In each state, a stratified random sample of 40 centers/schools/programs was selected from the list of all sites provided by state departments of education. The final sample consisted of 240 sites. Pre-K data collection for the SWEEP Study, designed to supplement the Multi-State Study, took place during the 2003–2004 school years in five states: New Jersey, Massachusetts, Texas, Wisconsin, and Washington. The 463 participating sites were recruited using the same method employed in the Multi-State Study. While not a random sample of classrooms, the 11 states included in these two companion studies account for approximately three-fourths of all children participating in state-funded pre-K programs in the U.S. (Early et al., 2005, p. 3).

In both the Multi-State and SWEEP studies, as in the current study, lead teachers responded to questionnaires regarding the classroom and their own backgrounds. Observational data collectors were all post-BA, full-time data collectors in the Multi-State and SWEEP studies and, as described above, were undergraduate students in child development and/or early education in the current study. Whereas, in the Multi-State and SWEEP studies four randomly selected children were observed in each classroom, we observed every child in the classroom, rotating among groups of four children at a time.

Because the NCEDL Multi-State and SWEEP studies included states that do not require that pre-K teachers have a BA degree or early childhood certification, a much smaller share of teachers had these credentials than was the case in Tulsa. Specifically, 69.6% of the teachers had a BA degree or higher and 47% had their early childhood certification. As a result, we restricted our cross-site comparisons to classrooms with a lead teacher who had both a BA degree and was certified to teach children ages 4 through third grade (to match the early childhood certification requirement in Tulsa). A total of 241 school-based pre-K classrooms and 25 Head Start classrooms from the multi-state NCEDL Multi-State/SWEEP sample met these criteria. The six Tulsa classrooms in which the lead teacher was not early childhood education certified were also excluded from the analyses involving comparisons across datasets.

3. Results

3.1. Data analysis

Both descriptive and inferential analyses were conducted. For the within-Tulsa descriptive data analyses, we report scores for the separate CLASS dimensions, including Student Engagement, as well as for the three composites: Emotional Support, Classroom Organization, and Instructional Support. For the comparisons with the multi-state data, we utilize a subset of the separate dimensions that were commonly collected in Tulsa and in the 11 participating states. Comparisons could not be made with any of the CLASS composites due to differences in the dimensions used to construct the composites in the Tulsa

and multi-state NCEDL studies. With regard to the CE-EAS, each of the five individual activity codes, as well as the Literacy Activities composite, was used in both the within-Tulsa analyses and the comparisons with the multi-state NCEDL data.

We first provide descriptive data on the TPS pre-K and CAP Head Start classrooms. We then examine differences in classroom climate and instructional time allocation in pre-K and Head Start classrooms in the Tulsa and multi-state NCEDL samples using ANOVAs and a Bonferroni correction given the large number of outcome measures. The Bonferroni-corrected threshold used to determine statistical significance ($p < .004$) was obtained by dividing the desired significance level (p-value of .05) by 14, the number of outcomes included in the analyses. Effect sizes (partial η^2) are reported for all main effects and interactions in the ANOVA results. These portray the proportion of variance in the dependent variable that is associated with the independent variable. In these comparative analyses, as noted above, only classrooms with BA-degreed teachers who had earned their early childhood certification are included.

Zero-order correlations among the classroom process (CLASS and CE-EAS) variables and the variables capturing teacher and classroom characteristics were examined as a prelude to the HLM regression analyses used to examine correlates of classroom processes in Tulsa. A Bonferroni correction was applied to the intercorrelations given the large number of measures. In this case, the Bonferroni-corrected threshold ($p < .0003$) was obtained by dividing the desired significance level (p-value of .05) by 171, the total number of correlation coefficients.

To examine sources of variation in classroom processes in the Tulsa pre-K programs, we used hierarchical linear modeling (HLM) analyses to take into account the nesting of classrooms within TPS schools and Head Start programs (Raudenbush & Bryk, 2002). The dependent variables are the three composites derived from the CLASS (Emotional Support, Classroom Organization and Instructional Support), the CLASS Student Engagement dimension, and the five CE-EAS items that capture the share of classroom time devoted to Math, Science, and other instructional activities, as well as the Literacy Activities composite. Independent variables are: whether the classroom is in a Head Start program (vs. a TPS school), whether the teacher has a BA in early childhood education (vs. a BA in something else), whether the teacher has an MA degree, the teacher's years of classroom experience, the teacher's undergraduate GPA, whether the classroom is half-day (vs. full-day), and the teacher's use of curricula (Direct Instruction, Integrated Thematic Instruction, and Creative Curriculum). Each of these variables is dichotomous, with the exception of the continuous measures of years (decades) of classroom teaching and GPA. The half-day vs. full-day comparisons reflect only on TPS classrooms given that the Head Start classrooms were all full-day.

Because 25 teachers did not grant us permission to obtain their undergraduate GPA (see above), missing data for this variable were imputed using STATA's ice program (Royston, 2005). We first generated five imputed values for each observation that was missing undergraduate GPA and then used the HLM program to run the two-level hierarchical model using plausible values analysis (used when only one variable is being imputed) for undergraduate GPA. Effect sizes (Cohen's d) were computed for all significant HLM results (Cohen, 1988). For indicator (dichotomous) variables, the effect size was calculated by dividing each coefficient (unstandardized) by the standard deviation (S.D.) of the outcome variable. Effect sizes for continuous variables were calculated by multiplying each coefficient by the S.D. of the predictor, then dividing by the S.D. of the outcome (see Belsky et al., 2007; NICHD ECCRN & Duncan, 2003).

3.2. Tulsa classrooms: descriptive findings

Table 1 presents the means and standard deviations for TPS and Head Start classrooms for all variables used in the study. TPS and Head Start classrooms tended to fall, on average, in the high to middle range of the CLASS scale (scores above 4.50) for most composites and dimensions. Nevertheless, substantial variation characterized these scores. For example, Emotional Support ranged from 3.25 to 6.80 in TPS classrooms ($M = 5.23$, S.D. = 0.57) and from 3.50 to 6.42 in Head Start classrooms ($M = 5.22$, S.D. = 0.78). The exception is the Instructional Support composite and its dimensions, which exhibited notably lower scores. The composite scores, for example, ranged from 1.40 to 5.94 in TPS classrooms ($M = 3.21$, S.D. = 0.93) and from 1.94 to 5.89 in Head Start classrooms ($M = 3.26$, S.D. = 0.94).

With regard to exposure to academic instruction, in both TPS and Head Start classrooms, teachers and children spent the largest proportion of time on literacy activities (M's = 30% in TPS and 24% in Head Start classrooms). They also spent over 10% of observed classroom time on activities that addressed Math, Science, Social Studies, and Aesthetics.

While about one-quarter of the TPS classrooms were half-day, all of the Head Start classrooms covered the full school day. Of the three curricula included in our analysis, TPS teachers tended to rely on Integrated Thematic Instruction and, to a lesser extent, Direct Instruction, whereas Head Start teachers relied more on the Creative Curriculum. The vast majority of the teachers who indicated using a specific curriculum reported that they used it daily or almost daily. Importantly, these curricula were not used in isolation from each other. For example, 18% of all teachers used both Direct Instruction and Integrated Thematic Instruction, 5% used both Direct Instruction and Creative Curriculum, 4% used both Integrated Thematic Instruction and the Creative Curriculum, and 3% used all three curricula. However, 38% of the teachers (33% in TPS and 54% in Head Start classes) did not use any of these three curricula.

A little over half (57%) of both TPS and Head Start teachers in our sample had received their BA in early childhood education. Only 16% of TPS teachers and 21% of Head Start teachers had an MA. On average, TPS teachers had 9 years and Head Start teachers had 7 years of teaching experience. Both TPS and Head Start teachers had an average undergraduate GPA of about 3.2, with a range from 2.32 to 3.98 for the total sample.

3.3. Comparisons between Tulsa and multi-state classroom samples

We began our cross-sample comparisons with a 2 (Sample: Tulsa/Multi-State) × 2 (Auspice: School-based/Head Start) MANOVA, performed on all 14 CLASS and CE-EAS outcome variables. Results indicated that there were significant main effects of Tulsa/Multi-State status, $F(14, 344) = 27.24$, $p < .001$, and School-based/Head Start status, $F(14, 344) = 4.68$, $p < .001$, and a significant Tulsa/Multi-State × School-based/Head Start interaction effect, $F(14, 344) = 4.03$, $p < .001$, on the combined dependent variables. We went on to conduct a series of 2 × 2 ANOVAs to investigate the main and interactive effects on each individual outcome.

Table 2 presents the findings from these ANOVAs. The analyses compare the classroom experiences of preschoolers in Tulsa to those of children in the multi-state classrooms, as well as the experiences of children attending school-based vs. Head Start 4-year-old classrooms in both samples (main effects). The interaction term captures whether differences found between the Tulsa and multi-state samples varied by auspice: school-based or Head Start.

The majority of findings reflect differences between classrooms in Tulsa and the multi-state comparison sample. Specifically, the classrooms in Tulsa were characterized by significantly higher scores on the CLASS dimensions of Productivity, $F(1, 358) = 22.05$, $p < .004$, Instructional Learning Formats, $F(1, 358) = 30.19$, $p < .004$, Concept Development, $F(1, 358) = 11.55$, $p < .004$, and Quality of Feedback, $F(1, 358) = 91.37$, $p < .004$. There were no significant differences between school-based and Head Start classrooms (across the two samples) on the CLASS dimensions, however, the significant interaction term for Concept Development, $F(1, 358) = 11.12$, $p < .004$, indicates that the difference between the Tulsa and multi-state classrooms was only significant for school-based classrooms (M's = 2.84 and 1.90 in school-based classrooms, M's = 2.42 and 2.41 in Head Start classrooms, for Tulsa and the multi-state samples, respectively). With the exception of the significant main effect of sample (Tulsa vs. multi-state) on Quality of Feedback, the effect sizes for these results are very modest.

With regard to instruction time (CE-EAS), again, the children in the Tulsa classrooms as compared to those in the multi-state sample received significantly more exposure to Literacy Activities, $F(1, 358) = 50.51$, $p < .004$, Math, $F(1, 358) = 27.77$, $p < .004$, Science, $F(1, 358) = 13.90$, $p < .004$, Social Studies, $F(1, 358) = 9.58$, $p < .004$, and Aesthetics, $F(1, 358) = 34.88$, $p < .004$ (see Table 2). Only time spent on Writing failed to show significant cross-sample differences. Main effects were also found for program auspice, such that teachers in school-based classrooms spent significantly more time than those in Head Start classrooms on Math activities, $F(1, 358) = 15.56$, $p < .004$, and significantly less time on Social Studies, $F(1, 358) = 19.29$, $p < .004$. The significant interaction for Social Studies, $F(1, 358) = 10.26$, $p < .004$, further revealed that the Tulsa Head Start teachers spent significantly more time on Social Studies activities than their counterparts in both the multi-state Head Start classrooms and the TPS classrooms. The significant interaction for Aesthetics, $F(1, 358) = 12.00$, $p < .004$, indicated that the Tulsa Head Start teachers spent significantly more time on Aesthetics than their counterparts in the multi-state Head Start classrooms. Again, the effect sizes for the instructional time results are very modest.

3.4. Intercorrelations among the classroom process and teacher and classroom characteristics measures

Table 3 presents the intercorrelations among the CLASS and CE-EAS variables, as well as the variables capturing teacher and classroom characteristics used in the HLM analyses for the within-Tulsa sample. The only salient pattern is one of positive associations among the CLASS scores, especially among the scores for Emotional Support, Classroom Organization, and Student Engagement. In contrast, the scores for the proportions of time spent engaged in CE-EAS instructional activities were not significantly intercorrelated, nor were any of the classroom process and teacher/classroom characteristics measures. This suggests that the two classroom observation instruments (CLASS and CE-EAS) captured distinct aspects of classroom processes, as expected.

3.5. Correlates of classroom processes within Tulsa

We next turn to the question of whether school auspice (Head Start or TPS), teacher characteristics, use of different curricula, or the half- or full-day schedule of the classroom were sources of variation in classroom processes. None of these independent variables was associated with the CLASS measures of classroom climate (see Table 4); however, the proportion of time spent in various instructional activities (CE-EAS observations) did show a few significant associations with whether or not the classroom was based in a Head Start program, the teacher's classroom experience and, at a marginal level, the teacher's use of specific curricula. Table 5 presents the HLM findings for the CE-EAS regarding Literacy and other instructional activities.

Head Start children spent significantly less time than their counterparts in TPS schools engaged in Math activities ($p < .05$, $d = -0.51$) and more time engaged in Social Studies activities ($p < .001$, $d = 1.21$). The magnitude of these auspice-related differences is not trivial. Auspice-related differences in time spent Writing, favoring the TPS teachers, were at the marginal level ($p < .10$, $d = -0.55$). With regard to teacher characteristics, only teaching experience emerged as a significant correlate of classroom processes. More experienced teachers devoted a significantly larger proportion of classroom time to Literacy Activities ($p < .05$, $d = 0.19$) and a borderline tendency to spend more time on Social Studies activities ($p < .10$, $d = 0.18$). The magnitude of these teacher experience effects is notably smaller than those found for program auspice. The only other correlate of time spent on classroom activities – and only at the borderline level of significance – involved the teacher's choice of curricula. Teachers who used Direct Instruction exclusively or in combination with other curricula devoted marginally

Table 2
Results of 2 × 2 analyses of variance examining classroom climate and instructional time allocation in school-based pre-K and Head Start classrooms in the Tulsa and multi-state samples (all teachers have BA degree and early childhood teaching certification).

Outcome variable	TPS pre-K classrooms ($n = 71$)		Multi-state school-based pre-K classrooms ($n = 241$)		Tulsa Head Start classrooms ($n = 25$)		Multi-state Head Start classrooms ($n = 25$)		Main effect of Tulsa/multi-state status ($n = 362$)		Main effect of school-based/Head Start status ($n = 362$)		Tulsa/multi-state[a] school-based/Head Start interaction ($n = 362$)	
	M	S.D.	*M*	S.D.	*M*	S.D.	*M*	S.D.	*F*	Partial η^2	*F*	Partial η^2	*F*	Partial η^2
CLASS dimensions														
Positive climate	5.04	0.82	5.25	0.87	4.87	0.87	5.26	0.66	4.94	0.01	0.32	0.00	0.46	0.00
Negative climate	1.36	0.57	1.46	0.70	1.53	0.78	1.34	0.38	0.22	0.00	0.06	0.00	2.01	0.01
Teacher sensitivity	4.82	0.67	4.73	0.91	4.61	0.88	4.79	0.64	0.12	0.00	0.32	0.00	1.01	0.00
Behavior management	5.00	0.83	5.02	0.99	4.33	1.31	4.81	0.94	2.65	0.01	8.54	0.02	2.23	0.01
Productivity	5.16	0.86	4.50	0.88	4.99	1.00	4.37	0.75	22.05[a]	0.06	1.23	0.00	0.03	0.00
Instructional learning formats	4.65	0.87	3.81	1.18	4.93	0.64	3.90	1.05	30.19[a]	0.08	1.17	0.00	0.29	0.00
Concept development	2.84	1.19	1.90	0.75	2.42	1.09	2.41	0.94	11.55[a]	0.03	0.11	0.00	11.12[a]	0.03
Quality of feedback	3.35	0.99	1.89	0.83	3.41	1.05	2.18	1.12	91.37[a]	0.20	1.57	0.00	0.67	0.00
CE-EAS dimensions (proportion of time spent on each activity)														
Literacy activities composite	.30	.13	.18	.08	.26	.09	.18	.09	50.51[a]	0.12	1.83	0.01	2.37	0.01
Writing	.03	.03	.02	.03	.03	.03	.02	.02	3.37	0.01	0.39	0.00	0.54	0.00
Engaged in math	.17	.10	.09	.06	.11	.05	.07	.04	27.77[a]	0.07	15.56[a]	0.04	3.66	0.01
Engaged in science	.17	.09	.11	.09	.18	.11	.14	.07	13.90[a]	0.04	1.55	0.00	0.57	0.00
Engaged in social studies	.13	.08	.13	.09	.24	.08	.15	.10	9.58[a]	0.03	19.29[a]	0.05	10.26[a]	0.03
Engaged in aesthetics	.18	.10	.15	.07	.24	.14	.12	.07	34.88[a]	0.09	0.74	0.00	12.00[a]	0.03

[a] *Note*: statistically significant using a Bonferroni-corrected threshold ($p < .004$).

Table 3
Intercorrelations among variables used in HLM analyses.

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. Emotional support composite	–																		
2. Classroom organization composite	0.75[a]	–																	
3. Instructional support composite	0.38[a]	0.28	–																
4. Student engagement	0.61[a]	0.80[a]	0.22	–															
5. Literacy activities composite	0.33	0.43[a]	0.07	0.25	–														
6. Proportion of time spent writing	−0.05	0.06	−0.01	0.01	0.26	–													
7. Proportion of time spent engaged in math	0.04	0.21	−0.11	0.09	0.31	−0.07	–												
8. Proportion of time spent engaged in science	0.14	0.23	−0.06	0.32	0.02	−0.14	0.10	–											
9. Proportion of time spent engaged in social studies	0.04	−0.05	0.03	−0.03	−0.06	−0.22	−0.16	0.19	–										
10. Proportion of time spent engaged in aesthetics	0.00	0.05	0.03	−0.02	−0.01	−0.04	0.14	0.06	0.08	–									
11. Half-day classroom	0.09	0.05	0.09	0.03	0.07	−0.02	0.12	0.17	−0.09	0.09	–								
12. Teacher uses direct instruction curriculum	−0.01	0.03	0.08	−0.01	0.17	−0.01	0.09	−0.05	−0.10	0.01	0.08	–							
13. Teacher uses integrated thematic instruction curriculum	0.01	0.01	0.07	0.04	0.06	−0.17	0.17	−0.06	−0.10	−0.10	0.09	0.29	–						
14. Teacher uses creative curriculum	−0.03	−0.07	−0.04	−0.01	−0.13	0.02	−0.19	0.04	0.24	0.15	−0.11	0.09	−0.14	–					
15. Teacher has BA in early childhood education	−0.04	−0.03	−0.12	0.01	0.17	0.08	0.04	0.05	−0.07	0.00	−0.01	0.03	−0.05	−0.05	–				
16. Teacher has MA	−0.07	−0.13	−0.10	−0.07	−0.10	−0.09	−0.02	0.05	0.03	0.04	−0.01	−0.04	0.09	0.12	−0.32	–			
17. Total decades as a classroom teacher	−0.04	0.08	0.01	0.03	0.14	−0.01	0.06	−0.14	0.14	−0.08	−0.07	−0.10	0.18	−0.09	−0.07	0.11	–		
18. Undergraduate GPA	0.03	0.10	0.12	0.03	0.13	−0.02	0.09	0.02	0.18	−0.17	−0.09	−0.25	−0.24	−0.14	0.06	−0.19	0.07	–	
19. Head Start classroom	−0.01	−0.10	0.02	−0.08	−0.22	−0.03	−0.30	0.02	0.49[a]	0.16	−0.27	−0.16	−0.34	0.46[a]	0.01	0.07	−0.09	0.07	–

[a] *Note*: statistically significant using a Bonferroni-corrected threshold ($p < .0003$).

Table 4
HLM analyses predicting CLASS scores from teacher and classroom characteristics (TPS and CAP Head Start classrooms combined; $n = 102$).

Variable	Emotional support composite *B* (S.E.)	Classroom organization composite *B* (S.E.)	Instructional support composite *B* (S.E.)	Student engagement score *B* (S.E.)
School-level variables				
Head Start classroom	0.04 (0.20)	−0.15 (0.20)	0.33 (0.21)	−0.16 (0.17)
Classroom-level variables				
Teacher has BA in early childhood education	−0.09 (0.12)	−0.11 (0.15)	−0.31 (0.20)	−0.01 (0.17)
Teacher has MA	−0.18 (0.18)	−0.30 (0.21)	−0.38 (0.28)	−0.18 (0.25)
Total decades as a teacher	−0.03 (0.12)	0.09 (0.10)	0.03 (0.14)	0.02 (0.10)
Teacher's undergraduate GPA	−0.08 (0.26)	−0.01 (0.28)	0.07 (0.36)	−0.04 (0.32)
Half-day classroom	0.17 (0.13)	0.08 (0.21)	0.32 (0.24)	0.04 (0.24)
Teacher uses Direct Instruction curriculum	−0.03 (0.14)	0.06 (0.15)	0.19 (0.22)	−0.06 (0.16)
Teacher uses Integrated Thematic Instruction curriculum	0.01 (0.14)	−0.05 (0.17)	0.16 (0.19)	0.04 (0.17)
Teacher uses Creative Curriculum	−0.09 (0.19)	−0.04 (0.29)	−0.22 (0.29)	0.09 (0.28)

Note: $^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

significantly more time to Literacy Activities ($p < .10$, $d = 0.44$), and teachers who used Integrated Thematic Instruction spent marginally less time on Writing ($p < .10$, $d = -0.36$) than teachers who did not use these curricula. The children's exposure to different instructional activities was not affected by the teachers' degrees or GPAs, nor by whether the classroom was full- or half-day.

4. Discussion

States have high aspirations for the investments they are making in pre-K education. The expectation is that high-quality, educationally focused programs for 4-year-old will reap benefits in improved school performance for all participating children. The pre-K program in Oklahoma, in particular, has received substantial attention from researchers and decision-makers alike. Not only is this program unusual in its universality, near exclusive reliance on school-based programs, relatively high teacher wages, and strict requirements for teacher qualifications and classroom size that apply to all participating programs, but it also has generated strong and replicated short-term results for children's achievement (Gormley & Gayer, 2005; Gormley et al., 2005, 2008) and, most recently, for social-emotional outcomes (Gormley, Phillips, Newmark, & Perper, 2009).

We first examined children's experiences with regard to classroom climate (Emotional and Instructional Support, and Classroom Organization) and exposure to academic instruction in Tulsa's pre-K classrooms. The quality of the classroom climates that children were exposed to in the pre-K and Head Start classrooms was generally quite high, albeit in the context of substantial variation, as would be expected based on other studies of classroom quality in pre-K (Early et al., 2007; Howes et al., 2008) and Head Start (U.S. Department of Health and Human Services, 2006). Students were also observed to be quite actively engaged in the learning activities provided by their teachers. It is, nevertheless, striking that the quality of Instructional Support was notably lower than the quality of Emotional Support in both public school (TPS) and Head Start classrooms. The classroom means for the CLASS Instructional Support scale fell in the low-middle range, whereas the Emotional Support means (as well as those for Classroom Organization) fell in the middle-high range. This finding is similar to that of other studies of pre-K using the CLASS (e.g., Early et al., 2006; Howes et al., 2008; Mashburn et al., 2008), as well as to observations of first-grade classrooms (NICHD ECCRN, 2002). Indeed, none of the TPS or Head Start classrooms in this study achieved either of the two highest scores on Instructional Support.

Relatively low Instructional Support scores may reflect actual differences between the two dimensions of classroom climate, such that it is more difficult for teachers to meet high standards for Instructional than for Emotional Support. Alternatively, this commonly observed disparity may suggest that the two scales are calibrated somewhat differently, such that a score of, for example, "6" on one scale is not comparable to a score of "6" on the other. In this context, it is important to note that the CLASS Instructional Support ratings were *not* significantly associated with observed time spent on specific instructional activities, nor with the CLASS composite for Classroom Organization or the Student Engagement dimension. Given that several studies have now found significant positive associations between Instructional Support and growth in child outcomes (Howes et al., 2008; Mashburn et al., 2008; NICHD ECCRN, 2002), our evidence lends further support to observations that greater time spent on instruction does not necessarily translate into children's exposure to high-quality instruction.

With regard to academic instruction, the most notable descriptive finding is the relatively large amount of time spent on Literacy Activities (consuming 24–30% of instructional time) as compared to time spent on Math, Science, or Social Studies. While this may be, in part, a function of the fact that the Literacy Activities score was a composite of several, separately coded activities, this result is consistent with other observational studies of pre-K classrooms (Howes et al., 2008), as well as of early elementary classrooms (NICHD ECCRN, 2002), and likely reflects the strong emphasis placed on literacy in early educational

Table 5
HLM analyses predicting CE-EAS scores from teacher and classroom characteristics (TPS and CAP Head Start classrooms combined; $n = 102$).

Variable	Literacy activities composite *B* (S.E.)	Proportion of time spent writing *B* (S.E.)	Proportion of time spent engaged in math *B* (S.E.)	Proportion of time spent engaged in science *B* (S.E.)	Proportion of time spent engaged in social studies *B* (S.E.)	Proportion of time spent engaged in aesthetics *B* (S.E.)
School-level variables						
Head Start classroom	−0.06 (0.04)	−0.02† (0.01)	−0.05* (0.02)	0.01 (0.03)	0.11*** (0.02)	0.01 (0.04)
Classroom-level variables						
Teacher has BA in early childhood education	0.04 (0.03)	0.01 (0.01)	0.02 (0.02)	0.01 (0.01)	−0.01 (0.02)	0.00 (0.02)
Teacher has MA	−0.01 (0.02)	−0.01 (0.01)	0.01 (0.02)	0.01 (0.03)	−0.01 (0.02)	−0.01 (0.03)
Total decades as a teacher	0.03* (0.01)	0.00 (0.00)	0.01 (0.01)	−0.02 (0.01)	0.02† (0.01)	0.00 (0.01)
Teacher's undergraduate GPA	0.03 (0.03)	−0.01 (0.01)	0.03 (0.03)	0.02 (0.03)	0.03 (0.03)	−0.03 (0.05)
Half-day classroom	0.01 (0.03)	−0.01 (0.01)	0.01 (0.03)	0.05 (0.03)	0.02 (0.02)	0.04 (0.03)
Teacher uses direct instruction curriculum	0.06† (0.03)	−0.00 (0.01)	0.02 (0.02)	−0.01 (0.02)	−0.00 (0.02)	0.00 (0.03)
Teacher uses integrated thematic instruction curriculum	−0.01 (0.03)	−0.01† (0.01)	0.02 (0.02)	−0.00 (0.02)	0.02 (0.02)	−0.02 (0.02)
Teacher uses creative curriculum	−0.02 (0.04)	0.01 (0.01)	−0.01 (0.02)	0.01 (0.03)	0.02 (0.03)	0.03 (0.05)

Note: ** $p < .01$.
† $p < .10$.
* $p < .05$.
*** $p < .001$.

contexts in the United States today. The similarity in educational backgrounds, classroom experience and undergraduate grade point averages across the TPS and Head Start samples of teachers is also noteworthy, although not surprising in light of the comparable requirements that apply to these two programs in Tulsa.

This overall portrait of Head Start classrooms in Tulsa is encouraging in light of prior evidence from the early 1990s that the quality of school-based programs exceeded the quality of Head Start programs (Goodson & Moss, 1992). Perhaps the national effort to upgrade the qualifications of Head Start teachers and to emphasize early literacy education, and, in Tulsa, the specific expectation that all Head Start teachers participating in the pre-K program have a Bachelor's degree, an early childhood teaching certification, and wages that are equivalent to the TPS pre-K and elementary teachers, has paid off in the classroom processes that affect children's pre-K experiences.

We next examined whether the Tulsa pre-K classrooms exhibited higher quality classroom climates and greater time on instructional activities as compared to a multi-state, and thus more typical, pre-K sample. We also compared pre-K classrooms operated under different auspices (school-based or Head Start) across the two samples. We found that the Tulsa and multi-state classrooms differed significantly on subscales that load on the Classroom Organization and Instructional Support composites of the CLASS, despite restricting our comparisons to classrooms led by teachers with a BA degree and an early childhood teaching certificate. The pertinent subscales capture effective time management, use of instructional techniques that maximize students' engagement and learning, reliance on lessons and activities that promote higher order thinking skills, and the provision of feedback that expands understanding. In fact, a recent RAND Corporation report found that California's preschools, on average, received a CLASS Instructional Support score that was 0.6 of a standard deviation lower than that of the average TPS pre-K classroom (Karoly, Ghosh-Dastidar, Zellman, Pelman, & Fernyhough, 2008). On subscales of the CLASS Emotional Support composite (e.g., positive climate and teacher sensitivity), however, the Tulsa pre-K classrooms did not differ from those in the multi-state sample. Program auspice was not associated with differences on any of the CLASS subscales.

The Tulsa pre-K teachers in both TPS and Head Start classrooms also devoted substantially more time to virtually every academic activity as compared to their counterparts in the multi-state sample. Across both samples, school-based teachers spent significantly more time on Math than did Head Start teachers who, in turn, spent more time on Social Studies. Taken together, this pattern of evidence from the Tulsa classrooms implies that it is possible to provide both relatively higher quality and larger amounts of academic instruction without sacrificing a positive climate, sensitivity to children's needs, and child-centered values. The findings also parallel the larger learning gains that have been found for the Tulsa pre-K program as compared to pre-K programs in other states (Gilliam & Zigler, 2004; Howes et al., 2008; Magnuson, Ruhm, & Waldfogel, 2007), and suggest that higher classroom quality and greater instructional time beget greater learning. This remains, however, to be tested directly.

While comparisons across research programs must be interpreted with caution, these findings also imply that policies requiring lead teachers in pre-K classrooms to have a BA degree and/or early childhood certification do not guarantee consistently higher quality classroom processes, nor greater student exposure to academic instruction (Early et al., 2007; Howes et al., 2008; Justice, Mashburn, Hamre, & Pianta, 2008; LoCasale-Crouch et al., 2007). Indeed, the Tulsa classrooms were superior on these features even when compared to classrooms in other states with similarly trained teachers. As such, these findings beg the question of why Tulsa pre-K teachers provide higher quality Instructional Support and Classroom Organization, and devote more classroom time to academic instruction than do college-educated teachers in other state pre-K programs. One possibility could be that the Tulsa classrooms have lower proportions of high-risk children as indexed by assessments of poverty. This does not appear to be the case, however, given that 63% of the children in TPS classrooms led by a BA-level teacher with an early childhood teaching certificate were eligible for a free lunch (130% of poverty) and 53% of the children in the comparable NCEDL Multi-State/SWEEP classrooms were from families living at or below 150% of poverty. The comparable figures for Head Start classrooms were 95% in Tulsa and 79% in the NCEDL Multi-State/SWEEP sample. Indeed, these numbers suggest that the Tulsa classrooms may have had a somewhat larger share of high-risk children.

One of the other prominent features of Oklahoma's pre-K program is that lead teachers are paid at the same levels (and receive the same benefits) as teachers who instruct older children, thus eliminating disincentives that might otherwise deter talented teachers who want to teach preschoolers from doing so. Given that pre-K teachers generally earn less than half of what the average elementary school teacher earns, this is a striking and somewhat unique feature of the Tulsa pre-K program. It is also the case that Oklahoma's average pre-K teacher is paid almost the same as the "average" wage earner in the state, which is also true of only two of the eleven states in the NCEDL Multi-State/SWEEP sample, making early childhood teaching a more attractive economic option than is the case in most other states.

The critical question for policy purposes concerns how to produce the kinds of pre-K classrooms that generate learning and social gains. Emerging evidence has attempted to get inside the "black box" of pre-K classrooms to address this question (Clifford et al., 2005; Early et al., 2006; Howes et al., 2008). Our contribution to this literature involved examining associations between the outcomes of classroom climate and instructional time allocation and the inputs of school auspice (TPS or Head Start); teachers' educational backgrounds, experience, and GPAs; their curricular choices; and whether the classroom was on a full- or half-day schedule.

None of these program or teacher characteristics predicted variation in the quality of the Instructional Support, Emotional Support, or Classroom Management that children experienced in Tulsa's pre-K programs. This is disappointing given that the CLASS provided our assessment of program quality and because the Instructional Support scale, in particular, has been associated with children's early learning in previous studies (Howes et al., 2008; Mashburn et al., 2008). It would be

extremely valuable to understand the classroom and teacher attributes that predict supportive teacher-student instructional interactions. To the extent that the answer involves teachers' general level of education, specific training in early education, and wages (see, for example, LoCasale-Crouch et al., 2007; Pianta et al., 2005), the lack of variation on these variables in Tulsa may have worked against finding significant predictors.

The teachers' allocation of instructional time was, however, modestly affected by some of the inputs we examined. Replicating the results from the comparisons with the multi-state sample, teachers in Head Start classrooms did not spend less time on Literacy Activities than did TPS teachers, but they did spend significantly less time on Math activities and more time on Social Studies, which included time in which the children were engaged in fantasy play. These differing distributions of classroom activities likely reflect both the more explicit educational focus of a public school system and the more comprehensive goals of Head Start.

In light of the low correlation between the quality (CLASS scores) and amount (CE-EAS scores) of instruction in the Tulsa sample, it remains to be seen whether the comparable levels of classroom Instructional and Emotional Support, or the differences in actual instructional time allocation across the TPS and Head Start classrooms, play the larger role in supporting academic and social-emotional outcomes at the end of the pre-K year for children in TPS and Head Start classrooms. Of course, it may also be the case that classroom climate and focus on instructional activities interact such that the quantity of instructional time is positively associated with early learning only under conditions of positive Emotional and Instructional Support. The national interest in fostering academic achievement in young children suggests that current efforts to enhance academic instruction within Head Start should be sustained, if not further improved. By the same token, the higher levels of Instructional Support and exposure to academic instruction within the Tulsa Head Start classrooms, as compared to the multi-state Head Start classrooms, suggest that the broader context within which Head Start operates in Tulsa provides an important platform upon which to build these types of initiatives.

Our efforts to identify other classroom and teacher characteristics that are associated with positive classroom processes were disappointing. Only the teachers' years of classroom experience showed a significant association, and only with the classroom time allocated to Literacy Activities. This finding corresponds to evidence from both the pre-K (Pianta et al., 2005) and elementary literatures, in which more experienced teachers provide higher quality classroom management and instruction and, in turn, produce stronger student achievement (Boyd et al., 2007; Clotfelter et al., 2006; Jacob, 2007). We examined whether this experience effect is attributable to novice teachers, but no such effects emerged among the Tulsa pre-K teachers. It appears, in this pre-K teaching sample, that experience exhibits a relatively linear, positive association with children's exposure to literacy instruction. In contrast, the teachers' specific educational backgrounds, their undergraduate grade point average, and their choice of curriculum did not play a significant role in the quality of classroom processes that we observed in the Tulsa pre-K classrooms. This corresponds to recent work in the child care, pre-K and education literatures calling into question the use of educational credentials as a proxy for teacher quality (Early et al., 2006, 2007; Jacob, 2007). Oklahoma's relatively stringent lead teacher requirements did not eliminate teachers with quite low GPAs (31% of the Tulsa pre-K teachers had GPAs below 3.00). The curriculum results, which were borderline at best, also confirm prior evidence that the teacher's choice of curricula is far less important to student achievement than is the skill with which she uses the curricula she selects (Cole et al., 2005; Ryder et al., 2006). Clearly, there is a tremendous need for continued efforts to decipher predictors of individual differences in teaching quality among BA-level teachers (or, for that matter, other teachers). Promising candidates include teacher motivation, self-efficacy, job commitment, and leadership support.

This study has several limitations. We did not collect data on assistant teachers and, as a result, are not able to assess their potentially important impact on classroom processes. Nor did we assess classroom composition with regard to the share of children with special needs or behavior problems, additional factors that could play a large role in affecting what happens inside the classroom door. Of specific relevance to the cross-study comparisons, we were unable to assess whether differences in the classrooms across sites other than those we examined, such as the composition of the children or the student–teacher ratios on the observational days, account for our pattern of results. It is also the case that the NCEDL multi-state data were collected several years prior to the Tulsa data collection, which may produce cohort effects, perhaps especially with regard to the growing national emphasis on literacy instruction. Finally, this is not an experimental study and thus cannot support causal conclusions.

In conclusion, this study yielded both interesting and disappointing results. Tulsa's pre-K programs, regardless of whether they are provided in TPS or Head Start classrooms, exceed their counterparts in other states in terms of both the amount and quality of instruction to which young children are exposed. This finding corresponds to the strong evidence that these programs have generated regarding their impact on children's cognitive and social-emotional development at kindergarten entry (Gormley et al., 2008, 2009). That this occurs even when the comparison classrooms are led by comparably highly educated teachers is a puzzle that remains to be solved. The current results provide modest evidence that the teachers' years of teaching experience is associated with how they distribute their instructional time in the classroom and no evidence that they are associated with classroom climate. In future research, promising dimensions to explore include teachers' mental health status, notably depression, enjoyment of and commitment to teaching preschoolers (including high-risk preschoolers), the quality of their undergraduate education, and the support they receive from their principals and other educational leaders in the community.

Perhaps the most policy-relevant conclusion from this study is its demonstration that a mixed-delivery system for pre-K *that brings all programs under the same umbrella of high-quality standards* can promote positive experiences for young children across program auspices. In this context, however, it is important to note that the overwhelming majority of Oklahoma's

4-year-old is being served by public schools, not by other service delivery organizations. Public schools, with their expertise in and commitment to education, may exert strong competitive pressure on other organizations that provide early childhood education services. A mixed-services delivery system in which public schools play a less conspicuous role might produce weaker academic benefits. In the end, it is the prediction of early learning and positive behavioral outcomes that provides the best metric for deciphering what matters in the daily lives of young children in pre-K settings.

References

Aaronson, D., Barrow, L., & Sander, W. (2003, February). *Teachers and student achievement in the Chicago public high schools (Working paper)*. Chicago, IL: Federal Reserve Bank of Chicago.

Barnett, W. S., Epstein, D. J., Friedman, A. H., Boyd, J. S., & Hustedt, J. T. (2008). *The state of preschool 2008: State preschool yearbook*. New Brunswick, NJ: National Institute for Early Education Research.

Belsky, J., Burchinal, M., McCartney, K., Vandell, D. L., Clarke-Stewart, K. A., Owen, M. T., et al. (2007). Are there long-term effects of early childcare? *Child Development*, *78*(2), 681–701.

Bowman, B., Donovan, M. S., & Burns, M. S. (Eds.). (2000). *Eager to learn: Educating our preschoolers*. Washington, DC: National Academy Press.

Boyd, D., Goldhaber, D., Lankford, H., & Wyckoff, J. (2007). The effect of certification and preparation on teacher quality. *The Future of Children*, *17*(1), 45–68.

Clements, D., Sarama, J., & DiBiase, A. (2003). *Engaging young children in mathematics: Standards for early childhood mathematics education*. Mahwah, NJ: Erlbaum.

Clifford, R. M., Barbarin, O., Chang, F., Early, D., Bryant, D., Howes, C., et al. (2005). What is pre-kindergarten? Characteristics of public pre-kindergarten programs. *Applied Developmental Science*, *9*(3), 126–143.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. D. (2006). *Teacher–student matching and the assessment of teacher effectiveness (NBER Working Paper #11936)*. Cambridge, MA: National Bureau of Economic Research.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cole, K. N., Mills, P. E., Jenkins, J. R., & Dale, P. S. (2005). Early intervention curricula and subsequent adolescent social development: A longitudinal examination. *Journal of Early Intervention*, *27*(2), 71–82.

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.

Currie, J. (2001). *A Fresh Start for Head Start? Brookings Institution Children's Roundtable Report*. Washington, DC: The Brookings Institution.

Early, D. M., Barbarin, O., Bryant, D., Burchinal, M., Chang, F., Clifford, R., et al. (2005, May). *Pre-kindergarten in eleven states: NCEDL's Multi-State Study of pre-kindergarten & study of state-wide early education programs (SWEEP)*. Chapel Hill, NC: National Center for Early Development and Learning.

Early, D. M., Bryant, D. M., Pianta, R. C., Clifford, R. M., Burchinal, M. R., Ritchie, S., et al. (2006). Are teachers' education, major, and credential related to classroom quality and children's academic gains in pre-kindergarten? *Early Childhood Research Quarterly*, *21*, 174–195.

Early, D. M., Maxwell, K. L., Burchinal, M., Bender, R. H., Ebanks, C., Henry, G. T., et al. (2007). Teachers' education, classroom quality, and young children's academic skills: Results from seven studies of preschool programs. *Child Development*, *78*(2), 558–580.

Ehrenberg, R. G., & Brewer, D. J. (1997). Evaluating the effect of teacher degree level on educational performance. In W. J. Fowler (Ed.), *Developments in school finance* (pp. 197–210). Washington, DC: National Center for Education Statistics.

Ehri, L. C., Nunes, S. R., Stahl, S. A., & Willows, D. M. (2001). Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's meta-analysis. *Review of Educational Research*, *71*, 393–447.

Ferguson, R. E. (1991). Paying for public education: New evidence on how and why money matters. *Harvard Journal on Legislation*, *28*, 465–498.

Gilliam, W., & Marchesseault, C. (2005, March). From capitols to classrooms. In *policies to practice: State-funded prekindergarten at the classroom level*. New Haven, CT: Yale Child Study Center.

Gilliam, W., & Zigler, E. (2001). A critical meta-analysis of all evaluations of state-funded preschool from 1977 to 1998: Implications for policy, service delivery and program evaluation. *Early Childhood Research Quarterly*, *15*, 441–473.

Gilliam, W., Zigler, E. (2004). *State efforts to evaluate the effects of pre-kindergarten: 1977–2003*. Unpublished manuscript. Yale University. New Haven, CT.

Goldhaber, D. (2007). Everyone's doing it, but what does teacher testing tell us about teacher effectiveness? *Journal of Human Resources*, *42*, 765–794.

Goodson, B., & Moss, M. (1992). *Analysis report: Predicting quality of early childhood classrooms. Observational study of early childhood programs*. Boston, MA: Abt Associates, Inc.

Gormley, W., & Gayer, T. (2005). Promoting school readiness in Oklahoma: An evaluation of Tulsa's pre-K program. *Journal of Human Resources*, *40*(3), 533–558.

Gormley, W., Gayer, T., Phillips, D., & Dawson, B. (2005). The effects of universal pre-K on cognitive development. *Developmental Psychology*, *41*(6), 872–884.

Gormley, W., Phillips, D., & Gayer, T. (2008). Preschool programs can boost school readiness. *Science*, *320*, 1723–1724.

Gormley, W., Phillips, D., Newmark, K., & Perper, K. (2009, April). *Social-emotional effects of early childhood education program in Tulsa*. Denver, CO: Paper presented at the biennial meeting of the Society for Research in Child Development.

Hamre, B. K., Mashburn, A. J., Pianta, R. C., LoCasale-Crouch, J., & La Paro, K. M. (n.d.). Classroom assessment scoring system: Technical appendix. Charlottesville, VA: University of Virginia.

Hamre, B. K., & Pianta, R. C. (2005). Can instructional and emotional support in the first grade classroom make a difference for children at risk of school failure? *Child Development*, *76*(5), 949–967.

Hanushek, E. A. (1971). Teacher characteristics and gains in student achievement: Estimation using microdata. *American Economic Review*, *60*, 280–288.

Hanushek, E. A., & Rivkin, S. G. (2004). How to improve the supply of high quality teachers. In D. Ravitch (Ed.), *Brookings papers on education policy 2004* (pp. 7–25). Washington, DC: Brookings Institution Press.

Haskins, R., & Sawhill, I. (2003). *The future of Head Start (Brookings Institution Policy Brief, Welfare Reform and Beyond #27)*. Washington, DC: The Brookings Institution.

Henry, G., Gordon, C., Mashburn, A., & Ponder, B. (2001). *Pre-K longitudinal study: Findings from the 1999–2000 school year*. Atlanta, GA: Georgia State University, Applied Research Center.

Hill, H. C. (2007). Learning in the teaching workforce. *The Future of Children*, *17*(1), 111–127.

Howes, C., Burchinal, M., Pianta, R., Bryant, D., Early, D., Clifford, R., et al. (2008). Ready to learn? Children's pre-academic achievement in pre-Kindergarten programs. *Early Childhood Research Quarterly*, *23*, 27–50.

Howes, C., & Richie, S. (2002). *A matter of trust: Connecting teachers and learners in the early childhood classroom*. New York, NY: Teachers College Press.

Jacob, B. A. (2007). The challenges of staffing urban schools with effective teachers. *The Future of Children*, *17*(1), 129–154.

Justice, L., Mashburn, A. J., Hamre, B. K., & Pianta, R. C. (2008). Quality of language and literacy instruction in preschool classrooms serving at-risk pupils. *Early Childhood Research Quarterly*, *23*, 51–68.

Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2006). *What does certification tell us about teacher effectiveness? Evidence from New York City (NBER Working Paper #12155)*. Cambridge, MA: National Bureau of Economic Research.

Karoly, L., Ghosh-Dastidar, B., Zellman, G., Pelman, M., & Fernyhough, L. (2008). *Prepared to learn: The nature and quality of early care and education for preschool-age children in California*. Santa Monica, CA: RAND Corporation.

La Paro, K., Pianta, R., Hamre, B., & Stuhlman, M. (2002). *Classroom Assessment Scoring System (CLASS). Pre-K version*. Charlottesville, VA: UVA.

La Paro, K., Pianta, R., & Stuhlman, M. (2004). The classroom assessment scoring system: Findings from the prekindergarten year. *The Elementary School Journal*, *104*, 409–426.
LoCasale-Crouch, J., Konold, T., Pianta, R., Howes, C., Burchinal, M., Bryant, D., et al. (2007). Observed classroom quality profiles in state-funded pre-kindergarten programs and associations with teacher, program, and classroom characteristics. *Early Childhood Research Quarterly*, *22*(1), 3–17.
Magnuson, K., Meyers, M., Ruhm, C., & Waldfogel, J. (2004). Inequality in preschool education and school readiness. *American Educational Research Journal*, *41*, 115–157.
Magnuson, K., Ruhm, C., & Waldfogel, J. (2007). Does prekindergarten improve school preparation and performance? *Economics of Education Review*, *26*, 33–51.
Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O., Bryant, D., 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.
NICHD Early Child Care Research Network. (2002). The relation of global first-grade classroom environment to structural classroom features and teacher and student behaviors. *The Elementary School Journal*, *102*(5), 367–387.
NICHD Early Child Care Research Network, & Duncan, G. J. (2003). Modeling the impacts of child care quality on children's preschool cognitive development. *Child Development*, *74*(5), 1454–1475.
Phillips, D., & Stipek, D. (1993). Early formal schooling: Are we promoting achievement or anxiety? *Applied and Preventive Psychology: Current Scientific Directions*, *2*, 141–150.
Pianta, R., Howes, C., Burchinal, M., Bryant, D., Clifford, R., Early, D., et al. (2005). Features of pre-kindergarten programs, classrooms, and teachers: Do they predict observed classroom quality and child–teacher interactions? *Applied Developmental Science*, *9*(3), 144–159.
Pianta, R. C., La Paro, K., & Hamre, B. (2008). *Classroom assessment scoring system manual*. Baltimore, MD: Paul H. Brookes.
Preschool Curriculum Evaluation Research Consortium. (2008). *Effects of preschool curriculum programs on school readiness. Report from the preschool curriculum evaluation research initiative*. Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education.
Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.
Resnick, G., & Zill, N. (2003, April). *Understanding quality in Head Start classrooms: The role of teacher and program-level factors*. Minneapolis, MN: Paper presented at the biennial meeting of the Society for Research in Child Development.
Ripple, C. H., Gilliam, W. S., Chanana, N., & Zigler, E. (1999). Will fifty cooks spoil the broth? The debate over entrusting Head Start to the states. *American Psychologist*, *54*(5), 327–343.
Ritchie, S., Howes, C., Kraft-Sayre, M., & Weiser, B. (2002). *Emerging academics snapshot*. Los Angeles: University of California.
Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, *73*(2), 417–458.
Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review*, *94*(2), 247–252.
Royston, P. (2005). Multiple imputation of missing values: Update. *The Stata Journal*, *5*(2), 188–201.
Ryder, R. J., Burton, J. L., & Silberg, A. (2006). Longitudinal study of direct instruction effects from first through third grades. *Journal of Educational Research*, *99*(3), 179–191.
Schweinhart, L. J., & Weikart, D. P. (1997). The high/scope preschool curriculum comparison study through age 23. *Early Childhood Research Quarterly*, *12*, 117–143.
Snow, C., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.
Stebbins, L., St. Pierre, R., Proper, E., Anderson, R., & Cerva, T. (1977). *Education as experimentation: A planned variation model, an evaluation of follow through* Cambridge, MA: Abt Associates.
Stipek, D., Byler, P. (2003). The early childhood classroom observation measure. Unpublished Manuscript. Stanford University.
Summers, A. A., & Wolfe, B. L. (1977). Do schools make a difference? *American Economic Review*, *67*, 639–652.
U.S. Department of Health and Human Services. (2006, December). *Head Start FACES findings: New research on Head Start outcomes and program quality*. Washington, DC: Author.
Wong, V., Cook, T., Barnett, W. S., & Jung, K. (2008). An effectiveness-based evaluation of five state pre-kindergarten programs. *Journal of Policy Analysis and Management*, *27*, 122–154.