Will public pre-K really close achievement gaps? Gaps in prekindergarten quality between students and across states

Forthcoming in *American Educational Research Journal*

Rachel Valentino
Stanford University

Contact information:
rsans130@gmail.com

Acknowledgements: This research was supported by grant award #R305B090016 from the Institute for Education Sciences, U.S. Department of Education to Stanford University. The author thanks Sean Reardon, Deborah Stipek, Kenji Hakuta, Claude Goldenberg, and Amanda Barnes Cook for their invaluable help and feedback on earlier versions of this work. The author would also like to thank Steve Barnett and the National Institute for Early Education Research for access to these data.

Abstract: Publicly funded pre-K is often touted as a means to narrow achievement gaps, but this goal is less likely to be achieved if poor/minority children do not, at a minimum, attend equal quality pre-K as their non-poor/non-minority peers. In this paper I find large "quality gaps" in public pre-K between poor/minority students and non-poor/non-minority students, ranging from 0.3 to 0.7 SD on a range of classroom observational measures. I also find that even after adjusting for several classroom characteristics, significant and sizable quality gaps remain. Finally, I find much between-state variation in gap magnitudes, and that state-level quality gaps are related to state-level residential segregation. These findings are particularly troubling if a goal of public pre-K is to minimize inequality.

Keywords: Early childhood education, public pre-K, classroom quality, gaps, segregation

By the time children first enter kindergarten, there are already large gaps in achievement between students of different racial, socioeconomic, and language backgrounds. More specifically, achievement gaps between Black and Hispanic children and their White peers, and between children from lower socioeconomic backgrounds and their higher socioeconomic counterparts, are about two thirds of a standard deviation at the start of kindergarten (Duncan & Magnuson, 2005; Loeb & Bassok, 2008 Reardon & Portilla, 2016; Reardon & Robinson, 2008)—the equivalent of about three years of learning in later grades. Further, gaps between dual language learners (DLL)[1] and their native-English-speaking peers are as large as a standard deviation in early elementary grades (Reardon & Galindo, 2009).

Publicly funded prekindergarten has long been touted as a means to bolster disadvantaged children's academic skills to reduce gaps in achievement before children enter kindergarten. The hope is that it will prevent "at-risk" children from falling further behind their peers in later grades (Camilli, Vargas, Ryan, & Barnett, 2010; Lee & Burkam, 2002; Pianta & Howes, 2009). There are three ways that public pre-K might help to narrow achievement gaps: (1) If pre-K quality experienced by poor/minority students is higher, on average, than that of non-poor/non-minority students; (2) If, despite equal levels of quality, poor/minority students benefit more from pre-K; and/or (3) If public pre-K is disproportionally attended or attended for more years by poor and minority students. This paper focuses primarily on investigating the first possibility, whether publicly funded pre-K is of equal or higher quality for poor/minority children compared to pre-K instruction received by their non-poor/non-minority peers.

Overwhelmingly, research indicates that the highest quality programs yield the largest benefits for children (e.g. LoCasale-Crouch et al., 2007; Mashburn et al., 2008); especially for children of disadvantaged backgrounds (Garces, Thomas, & Currie 2002; Magnuson, Ruhm, &

Waldfogel, 2007). While this is good news, as it indicates that there are levers that can be manipulated to improve the outcomes of "at-risk" children, it also indicates that if children do not have equal access to high-quality programs across racial, socioeconomic, and language groups, the goal of narrowing achievement gaps may be more difficult to achieve. Certainly, quality pre-K is not the only mechanism that might narrow achievement gaps. Educational quality in K–12, income inequality, residential segregation, access to extracurriculars, and a number of other factors also matter—but quality pre-K is one important mechanism that is increasingly politicized. Thus, understanding whether there are gaps in the quality of programs attended across groups and, if so, what factors may be driving such gaps is critical to increasing the likelihood that public pre-K will have its desired effect.

In this paper, I pursue four goals. First, I examine evidence-based indicators of the quality of preschool classroom process and structure to measure the average differences in quality of state-funded pre-K programs experienced by students of different racial, ethnic, socioeconomic, and language groups. Once I determine the magnitude of quality gaps, I then investigate *why* quality gaps are so large. In particular, I examine whether these gaps can be explained by structural program characteristics, differences in average initial academic and social skills among students in the classroom upon entry to the program, classroom demographic composition, teacher race, and teacher use of a language other than English in the class. Third, I examine whether quality gaps vary among states to consider whether some states are more successful at offering equitable pre-K quality across demographics. Finally, I consider whether state-level quality gap magnitudes can be explained by between-state differences in the rate of expansion of the state pre-K programs over time, in levels of public pre-K spending, and in

levels of residential segregation. This last question could have implications for how to improve policies in states where quality gaps are currently large.

Background Literature

What is High Quality Pre-K?

The quality of children's early childhood learning experiences has been measured in a number of ways, but the various measures of early childhood education (ECE) quality generally fall into two categories: *structural quality* and *classroom processes*.

Structural quality. Structural measures of quality include factors such as staff-child ratios, class-size, operation as full- versus half-day, classroom materials, and teacher credentials. Structural measures of quality are not often found to be direct predictors of children's academic and/or social-emotional outcomes, but are rather shown to *indirectly* predict child outcomes through their mediated impact on classroom process (Justice, Mashburn, Hamre, & Pianta, 2008; NICHD ECCRN, 2002). Structural measures of quality are commonly of interest to policy-makers because they are easily regulable and measurable.

Classroom process. Measures of classroom process focus on how learning happens in the classroom. Process generally captures teacher-child interactions, including teachers' sensitive and responsive caregiving, attunement to children's cognitive and emotional needs, use of strategies that scaffold children's learning, and use of open-ended questions and expansions to facilitate complex thought development among children. A large body of theory and research suggest that high-quality interactions between teachers and children are principal mechanisms that drive children's development (Mashburn et al., 2008; Morrison & Connor, 2002; NICHD ECCRN, 2002; Pianta, 2006; Pianta, Steinberg, & Rollins, 1995).

Figure 1 provides a conceptual framework for how state policies and other factors may influence classroom structure and composition, which in turn influence classroom processes, which subsequently impact student outcomes.

Insert Figure 1 about here

Are all quality measures created equal? States have become increasingly invested in finding ways to improve the quality of early childhood education programs (Hustedt & Barnett, 2011). Many states are using state-level Quality Rating Improvement Systems (QRIS), which attempt to improve preschool program quality by assessing programs using a series of quality measures and assigning programs a star rating that provide an aggregate approximation of performance across all measures (see Tout, Starr, Soli, Moodie, Kirby, & Boller, 2011). Quality metrics typically span structural and process measures of quality. Take for example the Early Childhood Environmental Rating Scale Revised Edition (ECERS-R) (Harms, Clifford, & Cryer, 2005)—one of the most widely used measures of early childhood quality, to date being used as the primary assessment of quality in tens of thousands of classrooms across numerous states, and included in many states' QRISs (Sabol & Pianta, 2014). The ECERS-R is designed to measure both classroom structure and process. For example, items in the activity subscale of the ECERS-R, which focuses on activities such as dramatic play, science, and math, measure *what* kinds of materials children have available to them and the proportion of the day allocated to those activities (structure), while other items in the language and reasoning subscale aim to measure *how* the teacher scaffolds communication among children (process).

But not all measures of classroom quality are created equal. Generally, measures of classroom process tend to have greater predictive power over student outcomes than structural measures (see Mashburn et al., 2008). Measures such as the ECERS-R have weak predictive

power over student outcomes—at best, studies have found that the language and interactions factor predicts children's academic, but not social-emotional, outcomes (see Burchinal, Howes, Pianta, & Barbarin, 2008; Clifford & Reszka, 2010; Peisner-Feinberg et al., 2001). On the other hand, process-based measures such as the Classroom Assessment Score System (CLASS) (Pianta, La Paro, & Hamre, 2008)—increasingly used to measure classroom process and teacher-child interactions in state pre-K and all Head Start programs—are a significant positive predictor of both children's academic and social-emotional outcomes (Mashburn et al., 2008).

A high-level overview of the predictive validity of measures analyzed in this paper can be found in Table 1. In this paper, I include both quality measures with strong predictive validity such as the CLASS, as well as weaker ones such as the ECERS-R, for two reasons. First, because states continue to use structural measures such as the ECERS-R as important benchmarks of quality, they are still relevant to the policy conversation. Second, because some research suggests that despite their limited *direct* predictive validity on children's development, measures such as the ECERS-R and other indicators of classroom structure may be important enablers of high quality classroom process and teacher-child interactions, which in turn impact child outcomes (NICHD, 2002; Nores & Barnett, 2014).

Insert Table 1 about here

Is the Definition of Process Quality Universal?

While classroom processes across measures are generally strong predictors of student outcomes (Mashburn et al., 2008) and typically provide information about what is happening in different classrooms, the use of measures to define "quality" is tricky, and attempts to do so are not always clear. Furthermore, what quality-measurement tools reward as "high quality" may impact children of different backgrounds in different ways.

For instance, one indicator of classroom process is the amount of time spent in free play (e.g. children's freedom to choose which play activity they will engage in and how they will play) versus direct instruction (e.g. teacher instructing the whole class during circle time), with more free play often seen as the gold standard. However, Chien et al. (2010) found that children in public pre-K who spent more time in free-choice activities realized significantly smaller gains on a range of academic outcomes than their peers in classrooms with more direct instruction. Although these short-term findings do not eliminate the possibility that free play is beneficial for other long-term outcomes, such as executive functioning, they do suggest that programs engaging in relatively more free play may be less likely to increase students' school readiness through academic knowledge and skills—the explicit goal of many targeted pre-K programs.

Definitions of quality may further be culturally defined (Baumrind, 1972; Chao, 2000; Fuller & Clarke, 1994; Howes, 2010; Kermani & Brenner, 2000; Ladson-Billings, 1995). In some communities (African American ones in particular), didactic and directive instruction is often seen as more desirable than student-directed exploratory play, while in others (mostly White, middle-class ones) scaffolded instruction is seen as the ideal (Pellegrini, Perlmutter, Galda, & Brody, 1990; Slaughter, 1987; Stipek, 2004). For this reason, it is not clear that classrooms serving students of different backgrounds *should* always look the same, as children may respond most positively to the style of caregiving they are most frequently exposed to and used to. Because of these ambiguities in what defines "high quality" pre-K, in the discussion I will elaborate on which quality gaps may represent the most unjust differences in quality.

The Link Between Pre-K and Child Outcomes, by Subgroup

When states fund pre-K, often an intention is to improve the future school performance of poor and minority children and, thereby, narrow the achievement gap. States presume to do this

by (a) expanding access to early childhood educational experiences for disadvantaged children and (b) providing programs that are high in quality, particularly for disadvantaged students.

Researchers have demonstrated that the largest preschool benefits result from the highest quality programs (Barnett, 2011; Burchinal, Vandergrift, Pianta, & Mashburn, 2010; Camilli et al., 2010). Furthermore, the overwhelming majority of rigorous early childhood evaluations are of programs serving children in poverty (Leak, Duncan, Li, Magnuson, Schindler, & Yoshikawa, 2010; Yoshikawa et al., 2013), and these studies generally find that high intensity programs (both in quantity and quality of instruction) yield large benefits for poor and/or minority children (see Campbell, Ramey, Pungello, Sparling, & Miller-Johnson, 2002; Schweinhart, Montie, Xiang, Barnett, Belfield, & Nores, 2005). Research is mixed on whether pre-K is *more* effective for disadvantaged children than their more advantaged peers, but at a minimum this research suggests that high quality pre-K programs are equally beneficial for poor and minority children as they are non-poor non-minority children (Gormley, 2008; Gormley, Gayer, Phillips, & Dawson, 2005; Magnuson, Lahaie, & Waldfogel, 2006; Weiland & Yoshikawa, 2013), and in a handful of cases may be more effective for the former group than the latter.

Given the above research, pre-K certainly has the potential to narrow achievement gaps, but mainly if (a) access exists across subgroups and/or (b) programs are high quality in nature. Less is known about whether the quality of state pre-K programs that disadvantaged children attend are, on average, high, and if so, whether they are as high as those attended by their more advantaged peers. Perhaps more importantly, if quality is unequal between these groups, it begs the question of what might be driving these inequalities.

Differential Access to Quality Pre-K by Student Backgrounds

Across preschool settings (including child care and Head Start), research suggests that poor and minority children are less likely to be in the highest quality and most stimulating programs relative to their White non-poor peers (Barnett, Carolan, & Johns, 2013; Early et al., 2010; see also Bassok, Fitzpatrick, Greenberg, & Loeb, 2013). Presumably, the purpose of public state pre-K is, in part, to narrow this gap in access to quality.

A handful of studies have considered whether state-funded pre-K programs serving higher proportions of poor and minority children are lower in quality than those serving more advantaged children. A 2010 case study of about 4,000 Georgia pre-K classrooms found that while higher minority lower-income communities had higher scores on structural measures of quality than did lower minority higher-income communities, they had the lowest scores on the CLASS (Bassok & Galdo, 2015). Using broadly defined high versus low classroom quality "profiles," another study of 238 classrooms in six states[2] found that programs scoring highest in quality served the lowest percentage of non-White and poor children (49%) compared to the lowest quality programs (73% and 65%, respectively). Finally, using data of the same 700 classrooms in 11 state pre-K programs analyzed in this paper, Chien et al. (2010) found that among four profiles of classroom instruction (free play, scaffolded, individual, and group instruction), higher proportions of Black, Hispanic, and poor children were enrolled in classrooms of the "individual instruction" profile (e.g. spending more time on didactic academic activities) than White non-poor children. The reverse was true of the other three profiles.

While those three studies shed light on the disparities in state pre-K quality and classroom process, I extend the Pianta et al. (2005) and Chien et al. (2010) studies in several important ways. First, the prior literature focused on broad profiles of quality, which is useful for thinking about equity of access overall, but makes it difficult to draw detailed policy

implications. I conduct analyses using a broader set of indicators of quality. Second, prior research focused on the proportion of children of various backgrounds enrolling in classrooms of different levels of quality. Instead, this paper's approach of presenting quality differences as standardized gaps at the student level has three advantages: (1) it allows for the comparison of gap magnitudes across measures on a uniform metric to understand whether quality gaps are a bigger challenge for some quality dimensions than others, (2) it allows one to compare the magnitude of quality gaps to the magnitude of achievement gaps, which are often computed in standard deviation units, and (3) it describes average differences in quality experiences of children of different subgroups rather than describing the average demographic composition of classrooms meeting some quality criteria. Third, the prior literature considered quality differences by income and race, with little attention to an ever-growing population of pre-K attendees—DLL students. This paper fills this gap. Fourth, while the prior literature spoke, in part, to the first question of this study (about whether there are differences in pre-K quality across student groups), they did not address any of the remaining questions—namely whether the size of "quality gaps" varies across states, which could have important policy implications if some states have larger gaps than others. Finally, the prior literature did not investigate which state- and classroom-level factors strongly predict quality gaps. This last piece is critical for understanding how to close quality gaps, and for crafting future policy to ensure that high quality pre-K is equitably distributed across groups.

Factors That Predict Classroom Process

There are a number of different factors that could lead to measured differences in pre-K process. The most obvious and perhaps salient include (1) structural differences, (2) differences

in students' skills prior to program entry, (3) differences in teacher and/or student race/ethnicity, and (4) state differences in pre-K policies and level of residential segregation.

(1) Differences in structural features of quality could lead to differences in process quality if, for example, programs serving higher proportions of poor and minority children are those employing teachers with less training and/or skills. If pre-K is anything like K–12, programs in the poorest neighborhoods may have the most difficulty attracting and retaining good teachers (Lankford, Loeb, & Wyckoff, 2002). (2) Differences in students' skills prior to program entry could also drive differences in classroom process if (a) classrooms serving more disadvantaged children are more difficult to manage (because of greater behavioral challenges), (b) teachers target instruction toward more basic skills to catch disadvantaged students up academically, or (c) teachers of lower credentials/skills are differentially sorted into these classrooms within schools. (3) Differences in teacher race across programs could drive differences in classroom process if teachers have different cultural ideals about child-rearing and thus engage in practices not typically rated high by conventional measures of pre-K process (Chao, 2000; Kermani & Brenner, 2000; Slaughter, 1987). Student race could affect process if teachers adapt their teaching styles to be observant of community definitions of appropriate caregiving (see Fuller & Clarke, 1994; Howes, 2010; Slaughter, 1987; Stipek, 2004). Finally, (4) states could influence differences in classroom processes either through state policies (e.g. state-mandated pre-K funding per child), which in turn may influence structural features of quality and thus classroom processes, or through their lack of oversight or ability to change characteristics of their states.

Virtually all of the existing literature on which factors predict quality have considered the degree to which structural indicators of quality predict classroom process. These studies tend to

indicate that there is some, albeit weak, relationship between factors such as teacher credentials and instructional quality (LoCasale-Crouch et al., 2007; NICHD, 2002; Nores & Barnett, 2014; Pianta et al., 2005), but not higher order teacher skills like language modeling (Justice et al., 2008).

State Pre-K Policy and Between-State Variation in Quality

There is much between-state variation in how state pre-K is regulated, which is likely to have implications for the quality of pre-K classroom process (see Figure 1; see also Justice et al., 2008; NICHD ECCRN, 2002). For example, to date just over half of all state pre-K programs require lead teachers to have a BA, and 60% require regular site visits (Barnett et al., 2013). One study found that there was more between-state variation in pre-K quality than there was within-state variation (Pianta et al., 2005). If this is true, then considering which state-level features are the strongest predictors of high quality pre-K seems critical for reforming state policy as a means to improve quality.

There is also much between-state variation in state pre-K spending per child, the rate at which states have increased the proportion of children served in pre-K over time, and the state's level of residential segregation (see Appendix tables available upon request; see also, Barnett et al., 2013). Any of these factors could explain quality gap magnitudes if they lead to selection of the best pre-K teachers into the most affluent pre-K programs. One hypothesis is that states that pay more per child might be able to recruit pre-K teachers from a broader, more highly qualified labor force to fill positions in densely populated poor and minority neighborhoods. Another is that states that expanded their pre-K programs the quickest will experience larger quality gaps, because as more pre-K teaching positions become available within a state, the best pre-K teachers sort into higher SES programs. Similarly, K–12 research has found that teachers have

preferences to teach in schools with large numbers of White, high-ability students (Boyd, Lankford, Loeb, Ronfeldt, & Wyckoff, 2011; Jackson, 2009), so states with high levels of residential segregation may also be those with the largest pre-K quality gaps because of limited access to high quality teachers. While much is known about differences across states in pre-K regulations, less is known about differences across states in process quality experienced by children of different racial, language, and socioeconomic backgrounds.

The Current Study: Data and Methods

Sample

Data for this study came from two main data sets, the State-Wide Early Education Programs Study (SWEEPS) and the National Center for Early Development and Learning (NCEDL) Multi-State Study of Pre-Kindergarten. Data for the studies were collected in a total of 11 states—five (Massachusetts, New Jersey, Texas, Washington, and Wisconsin) in 2003–2004 and six (California, Georgia, Illinois, Kentucky, New York, and Ohio) in 2001–2002, respectively. These two studies implemented the same measures and were designed to combine data sets. The states were selected to represent those that had committed significant resources to state pre-K initiatives and that had been operating for several years. At the time of data collection, 80% of all children in the United States participating in state-funded pre-K were enrolled in one of these states, and 83% of all state dollars spent on pre-K were spent in these states (Barnett, Robin, Hustedt, & Shulman, 2003).[3] The SWEEP study involved a stratified random sample of 100 sites in each of the five states, and the multi-state study involved a stratified random sampling of 40 state pre-K sites within each of the six states. Across both studies, one classroom was randomly selected per pre-K site. A total of 647 classrooms with

12,334 pre-K students (three and four years old) make up the final sample, which represents 90% of all sampled classrooms that had complete data on all classroom quality measures.

Classroom Quality Measures

Quality in this study was primarily measured through direct classroom observation (ECERS-R, CLASS, and the emerging academics Snapshot—more details below). A handful of measures of classroom structure were also measured using teacher surveys. Finally, one measure of teacher beliefs about child-rearing was also included. While this measure is not typically considered a measure of pre-K quality, it is included because teacher beliefs may play a critical role in determining how teachers facilitate classroom process. Means and standard deviations of all measures are presented in Table 2. The pre-K quality measures studied in this paper (described in more detail in the section that follows) were selected on three main criteria. First, to represent the most policy-relevant and widely used existing measures (e.g. ECERS-R and structural measures). Second, to include measures that predict a range of children's developmental outcomes (e.g. Snapshot for academic outcomes and CLASS for social-emotional ones). Third, to represent those measures with the strongest predictive validity for a range of children's outcomes (e.g. the CLASS).

Survey measures of structural quality. I use four main measures of structural quality in this paper. These include staff-child ratio, class size, teachers' years of experience with preschool-aged children, and teachers' years of education. These measures were assessed through director surveys for the former two questions and teacher surveys for the latter two. I include these measures because of their policy-relevance and highly regulated nature, despite mixed evidence on their direct relation to children's developmental outcomes (Early et al., 2007).

Classroom Assessment Scoring System (CLASS). The pre-K version of the CLASS, a measure of classroom process, was used in this paper (Pianta, La Paro, et al., 2008). Pianta et al. (2005) conducted a factor analysis of this study's version of the CLASS. This analysis yielded two factors of process quality that I use in this paper: *Emotional Climate* (derived from items Positive Climate, Negative Climate, Teacher sensitivity, Over Control, and Behavior Management) and *Instructional Climate* (derived from items Productivity, Concept Development, Learning Formats, and Quality of Feedback). Each of these two factors has been found to significantly predict children's gains in academic and social-emotional outcomes, respectively (Mashburn et al., 2008). Each of the nine items[4] were scored approximately eight times per classroom and were rated on a 7-point scale. Internal consistencies for the emotional and instructional climate scales were 0.86 and 0.78, respectively. All observers were trained prior to data collection and reliabilities on the CLASS were calibrated against videotape master codes or the codes of the measure's authors when establishing reliability through live classroom observations. Classroom assessors' mean weighted kappa[5] on their final reliability test was 0.73, and 93% of data collector responses were within one point of master codes.

Early Childhood Environmental Rating Scale-Revise (ECERS-R). The ECERS-R, a 43-item measure of structural and process classroom quality was also used. Each item is scored on a 7-point Likert scale. The instrument evaluates programs across seven domains: Space and Furnishings; Personal Care Routines; Language and Reasoning; Activities; Interactions; Program Structure; and Provisions for Parents and Staff. Consistent with prior research (e.g. Peiser-Feinberg et al., 2001), the original study researchers found two ECERS-R factors (see Pianta et al., 2005) that are used in the current analysis: Language and Interactions and Provisions for Learning. The former is a composite of indicators measuring factors like staff-child interactions,

discipline, and using language to develop reasoning skills. The Provisions for Learning composite measures structural factors such as materials. At best, only the language and interactions factor has been found to moderately predict children's academic gains (Burchinal et al., 2008; Clifford & Reszka, 2010; Peisner-Feinberg et al., 2001). However, I include the ECERS-R in this paper because of its wide usage across states as a key policy metric to measure quality (e.g. in QRIS systems; Sabol & Pianta, 2014).

Observers' reliability on the ECERS-R was determined prior to data collection against a master coder's scores during a practice classroom visit. Assessors' mean weighted kappa with the master coder was 0.74, and 87% of assessors' responses were within one scale point of the trainer's code levels. These numbers are consistent with other studies (e.g. Peisner-Feinberg et al., 2001).

Emerging Academics Snapshot (Snapshot). The Snapshot is a pre-K observation measure that captures moment-by-moment experiences of children in classrooms (Ritchie, Howes, Kraft-Sayre, & Weiser, 2001). The Snapshot is an important measure to include in this paper because it captures both *what* children experience in the classroom (e.g. share of the day spent in science), and *how* those experiences occur (e.g. free play, group instruction, etc.)—both inputs of which have been linked to children's academic outcomes (Chien et al., 2010). The Snapshot is conducted by observing four randomly selected children within each classroom for 20-second observation periods followed by 40-second coding periods. Each of the four children are observed in succession before returning to observe the first child again. The cycle continues for five iterations, after which the observer pauses to code other measures before returning to the Snapshot to complete additional cycles. On average, children were observed across 50 one minute cycles (*sd* = 22.4). The Snapshot codes cover three domains and are measured as present

or absent during each 20-second observational period. Codes for each domain and item were averaged across children within classrooms to obtain classroom-level measures of the average proportion of time children spent engaged in various activities and settings.

The Snapshot consists of three domains, indicators of which were coded if present during each observation interval: (1) *Setting*, (2) *Activity*, and (3) *Teacher-Child Interactions*. The (1) *Setting* domain captures three mutually exclusive categories, *free choice*, *teacher-assigned*, or *meals/routines*. The setting was also coded as *whole-group* or *small-group*. The (2) *Activity* domain captures any of 11 learning activity codes, which were not mutually exclusive (e.g. a child could have been read to about science). Sample activities are *Literacy* or *Math*. Finally, (3) *Teacher-Child Interactions* are coded in two ways. First, each observation interval codes whether teacher-child interactions (if present) were *Scaffolded* or *Didactic*. Interactions were then coded for whether they were present at all (*None*), or were *Minimal*, *Routine*, *Simple*, or *Elaborated* in nature. These were mutually exclusively coded, whereas scaffolded and didactic were not. Finally, the coder measures whether teachers were *Distracted*.

Teacher beliefs about child-rearing. The Modernity Scale (Schaefer & Edgerton, 1985) measured teachers' beliefs about child-rearing to differentiate teachers with more traditional adult-centered (authoritarian) beliefs from those with more child-centered (authoritative) beliefs. Prior research has shown that in child-care homes, caregiver attitudes about children and child-rearing significantly predict both classroom quality and the presence of behavior problems in young children (Clarke-Stewart, Lowe Vandell, Burchinal, O'Brien, & McCartney, 2002). Authoritative beliefs are generally associated with higher quality. Child-centered belief items were reverse coded to derive average scores on the measure. Teachers holding a more adult-centered view would agree with items like, "Children must be carefully trained early in life or

their natural impulses make them unmanageable," and "Children should always obey the teacher." Teachers with more child-centered beliefs would more strongly agree with statements such as "Children should be able to disagree with their parents if they feel their ideas are better." The measure is a 15-item Likert questionnaire ranging from 1 (do not agree) to 5 (strongly agree). Chronbach's alpha for this scale is 0.84 (Schaefer & Edgerton, 1985).

Explanatory Variables Used in Research Questions 2 Through 4

Family questionnaire. Questionnaires were sent home with all children (not only the four randomly selected) in each classroom. This questionnaire collected information about the ethnicity, gender, income-to-needs ratio, maternal education, and language status of children. This enabled me to precisely estimate the racial/ethnic and socioeconomic information of children in each of the 647 classrooms in the sample, representing 12,334 children.

Classroom-level explanatory variables. In addition to the above described variables, I used other variables to explain quality gaps. First, each average classroom's student performance (the average score among the four randomly sampled children per classroom) was computed to consider whether the average skills of students in the class might explain classroom quality. While using just four randomly sampled students to draw inferences about the average skills of all students in the classroom may generate less precise estimates due to small sample sizes, they should not be biased in aggregate analyses across classrooms. I used three variables. (1) *The Peabody Picture Vocabulary Test- 3rd edition* (PPVT-III) to measure receptive vocabulary (Dunn & Dunn, 1997). (2) The *Woodcock-Johnson Psycho-Educational Battery* (WJ) (Woodcock, McGrew, & Mather, 2001) Applied Problems (AP) subtest to assess emergent math reasoning and problem-solving skills. Spanish versions of both tests were administered to DLLs whose first language was Spanish. This score was used in place of the English score when applicable. (3)

The *Teacher-Child Rating Scale* (TCRS) (Hightower et al., 1986) social competence subscale was completed by teachers. This subscale consists of 20 items rated on a 5-point scale. I used additional measures of structural quality to explain process quality gap magnitudes. These measures include whether the class was full-day, hours/week students spent in class, and the teacher's wage on a per-hour basis adjusted to 2014 dollars, in addition to the structural measures used to compute quality gaps. Finally, teacher race and whether a teacher spoke a language other than English in the classroom were used to predict quality gaps. Of the 37.5% of teachers who spoke a language other than English in the classroom, 91% spoke Spanish.

I standardized raw scores on all continuous measures relative to the full sample in all states, to reflect the classroom's or children's performance relative to others in the study population.

State-level explanatory variables. I constructed three sets of variables to answer the final question about which state-level factors are correlated with state-level quality gaps. First, I collected state pre-K spending per child in the study year from the State of Preschool Yearbooks[6] (Barnett et al., 2003). Second, I calculated the rate of pre-K expansion from the Current Population Survey (CPS) and Kids Count data. CPS was used to construct a variable indicating the proportion of age-eligible children enrolled in public preschool[7] by year and state. Kids Count data allowed me to disaggregate the number of preschoolers enrolled in public school from those enrolled in Head Start. Finally, I used Census tract-level data from the American Community Survey (ACS) to compute the Information Theory Index (Theil's H), a measure of state-level residential segregation. The index increases from 0 to 1, where 0 represents complete integration and 1 represents complete segregation.

Sample of children. Sample students in the current paper are enrolled in state pre-K, which is often targeted to low-income children and could raise concerns about the external validity of findings. The average annual family income of the non-poor (above 150% of the poverty line) children in the sample is $58,610, compared to $29,790 for poor children. The income distribution in the sample is presented in Figure 1, with a dashed line indicating the poor/non-poor cut-off used for an average-size family in the current paper. The median family income in the United States is captured by the solid line. Generally, this distribution looks similar to the income distribution of the overall population in the U.S. with two exceptions—(1) there is a slightly higher proportion of poor children in the current sample, and (2) there is an income ceiling of just over $110,000 per year in the non-poor sample. While not entirely representative of the population, this income ceiling likely indicates that any poor-non-poor quality gaps presented are conservative or *underestimates* of the reality of such gaps.

Analytic Strategy

I analyzed data using linear regressions that adjusted for the survey design structure of the data, including primary sampling units (PSUs) or clusters within which classrooms were randomly selected. Probability weights allowed me to account for the number of classrooms in the population each observation represents.[8] In order to calculate quality gaps at the student-level, I used two variables to expand the classroom-level data set: proportion of students of each ethnic or socioeconomic background in each class and class size. The product of these variables yielded the number of students of each background in each class, which allowed me to create a long data set with dummy variables for whether student *i* in classroom *c* was Black, Hispanic, poor, and/or DLL. This approach enabled me to capture all students in each classroom (totaling 12,334 students), rather than relying on the four randomly sampled children within each

classroom to compute quality gaps. This method is also the most straightforward for calculating gaps, as it allows one to incorporate analytic weights through the expansion process, and then adjust for probability weights during regression models. I also estimated jackknife standard errors[9] in all regression models (Kolenikov, 2010).

It is important to note that the expanded analytic approach is mathematically the same as using an analytic weighting approach (Gould, 1999). I conducted analyses on a non-expanded data set as a specification check to ensure that the expansion method did not produce artificially small standard errors due to increased sample size. Here, Q is a classroom quality measure. To examine the average difference in Q between the classrooms of students of group A and B, I first standardized Q, and then computed $\bar{Q}_A - \bar{Q}_B$, where $\bar{Q}_A = \frac{\sum_c w_c t_c \pi_{Ac} Q_c}{\sum_c w_c t_c \pi_{Ac}}$, where c indexes classrooms, w is a classroom weight, t is the total enrollment in the classroom c, and π_{AC} is the proportion of students in classroom C who are in group A. The same is computed for group B and the two values are differenced. Both methods produced identical gap and standard error estimates.

To answer question one about gaps in pre-K classroom quality, I regressed a standardized (mean 0) version of each quality measure (Q_{ic}) for each student *i* in each classroom *c* on a dummy variable, G_{ic}, to represent student *i*'s group (e.g. race).[10] Quality gaps are standardized relative to the entire sample of students. I ran four separate regressions for each quality outcome (to capture differences between Black and White, Hispanic and White, poor and non-poor, and DLL and non-DLL groups, each separately), using a subpopulation option[11] to yield a constant sample across models estimated.

To answer question two, about which factors explain quality gap magnitudes, a model of the following form was estimated:

$$Q_{ics} = \beta_0 + \beta_1 \mathrm{G_{ics}} + \mathrm{X}_c + \Gamma_s + e_{ics}$$

Where Q_{ics} is the standardized quality experienced by student *i* in classroom *c* in state *s*. β_0 captures the standardized classroom quality experienced by the reference group (e.g. White students). β_1 captures the standardized difference between the group of interest and the reference group (e.g. Black versus White) in the classroom quality experienced, conditional on controls. For measures where higher scores represent higher quality, a negative coefficient indicates that poor and minority students receive lower-than-average quality pre-K than their respective counterparts. A handful of measures are coded in the opposite way by design. For example, higher scores on the Modernity Scale represent more traditional/authoritarian child-rearing views, so a positive coefficient on β_1 still represents an undesirable gap. A similar principle applies to class size.

X_c captures a vector of classroom-level controls used to explain quality. The vector includes four categories of variables: (1) structural characteristics of classrooms, to examine whether, for instance, variation across classrooms and student subgroups in teacher credentials and class-size explains the size of gaps, (2) average academic and social skills of students in the classroom upon fall entry, to explore the possibility that if the initial skills of students are low, teachers may target teaching to their students' ability, (3) demographic classroom composition, to consider the possibility that teachers teach differently in classrooms that are densely populated by one group over another, and (4) teacher race and a dummy for whether teachers speak a language other than English in the class, to examine whether these characteristics are tied to different practices. I add each set of controls to the model separately in four iterative regressions, followed by two additional models: one with state fixed effects, Γ_s, and one with all controls plus Γ_s to estimate within-state gaps and control for all unobservable idiosyncratic state factors

that influence classroom quality such as state pre-K policy. Observing the degree to which β_1 changes in magnitude and significance after the inclusion of these variables indicates how much of the relationship between student race and classroom quality is explained by other factors.

To answer question three, whether there is between-state variation in quality gaps, I estimate a model of the following form:

$$Q_{ics} = \Gamma_s + \Gamma_s * G_{ic} + e_{ics}$$

Where $\Gamma_s * G_{ic}$ represents a vector of state-by-subgroup dummy variables to estimate quality gaps within each state.

Finally, because of the limited number of states, I consider question four, regarding which state-level factors are correlated with quality gaps, qualitatively through figures. Due to data-use agreements, state names have been redacted from figures answering questions three and four. States are instead assigned a state number that is constant across figures.

Results

Question 1: Magnitude of Standardized Quality Gaps

Results for research question one, regarding the magnitude of standardized White-Black, White-Hispanic, poor-non-poor, and DLL-non-DLL pre-K quality gaps are presented in Table 3.

*** Insert Table 3 ***

Results indicate that quality gaps are large and statistically significant on most measures, generally ranging from 0.3 to 0.7 standard deviations (*sd*),[12] with magnitudes that tend to mirror the magnitude of achievement gaps when children first enter school (Duncan & Magnuson, 2005; Reardon & Portilla, 2016). Quality gaps tend to be largest on measures of emotional and instructional climate from the CLASS—estimated at around 0.4 to 0.7 *sd* depending on the subgroup—which equates to about 0.5 points out of 7 on this measure. It is noteworthy that these

CLASS factors are also the most highly predictive of student gains in social-emotional and academic skills, respectively. Finally, these gaps tend to be largest between Black and White students, especially on the CLASS (overall CLASS *gap*= -0.659 *sd*), about twice the size of Hispanic-White and poor-non-poor gaps on this measure.

Further, the Snapshot reveals some interesting patterns. Black, Hispanic, poor, and DLL students are all significantly less likely than their White, non-poor, and non-DLL peers to be engaged in free-choice activities (*gap*= -0.396 to -0.582 *sd*). These gaps equate to 6.2% to 9.1% less share of the school day spent in free-choice activities. In contrast, and perhaps instead, they are significantly more likely to be engaged in individual time (e.g. assigned to work individually on worksheets: *gap*= 0.348 to 0.423 *sd,* or 2.3% to 2.6% greater share of the day). The same pattern is generally apparent for didactic (*gap*= 0.212 to 0.527 *sd,* or 2.6% to 6.5% greater share of the day) versus scaffolded instruction (*gap*= -0.199 to -0.301 *sd,* or 1.4% to 1.9% less share of the day). Finally, the potential academic tradeoff between time allocated to letters/sounds and away from science is apparent. Black, Hispanic, poor, and DLL students spend significantly more time on activities related to learning letters/sounds (*gap*= 0.182 to 0.424 *sd,* or 0.7% to 1.6% greater share of the day) and significantly less time engaged in science activities (*gap*= -0.259 to -0.469 *sd*, or 1.4% to 2.6% less share of the day) than their non-minority peers, though the effect was not significant for DLL students. While these gap magnitudes, in terms of proportion of the overall day, seem small, when considering that on average pre-K classrooms spend only 4.1% of the day on letters/sounds and 7.2% of the day on science, for example,[13] these gaps are quite sizable.

Gaps on structural measures of quality are less salient than the others. Still, a few significant gaps stand out. First, the class sizes of Hispanic students are significantly larger than

those of their White peers (*gap*= 0.40).[14] In other words, Hispanic children have an additional 1.6 students in their classrooms, on average, than White students. Second, Hispanic, poor, and DLL students all have teachers with fewer years of experience than their respective counterparts (*gap*= -0.138 to -0.396, or 1 to 2.7 fewer years of experience in pre-K). Finally, it is noteworthy that DLLs have teachers with *more* years of education than do non-DLLs (*gap*= 0.219 *sd*, or approximately 0.4 more years of education). This may be because some states require additional credentials among teachers serving DLLs (e.g. bilingual or ESL certification).

There are also significant traditional child-rearing views gaps among all groups, such that minority and poor students are more likely to have teachers who endorse more adult-centered authoritarian beliefs than are their non-minority and non-poor peers.

Question 2: Classroom Factors That Explain Gap Magnitudes

Results for question two regarding what classroom factors explain gap magnitudes are presented in Figures 3a and 3b. Gaps included in this second analysis are limited to the largest and most significant quality gaps, and those that are generally the most significant predictors of student outcomes. The first figure in the series demonstrates gaps, unadjusted for any classroom-level controls. Each figure thereafter adjusts for a separate set of controls. Solid bars indicate gaps that are statistically distinguishable from zero. Hollow bars indicate gaps that are not statistically significant (or that cease to be statistically significant after adjusting for controls).

Insert Figures 3a & 3b

First, with a few exceptions, the set of structural classroom indicators[15] explain virtually none of the quality gap magnitudes on process quality. For example, the overall CLASS gaps remain equally large and significant before and after controlling for these factors. The structural

indicators are often jointly significant predictors of quality, indicating that structural factors covary with process quality, but do not vary much between groups.

Second, the set of average student ability variables[16] and classroom racial/socioeconomic composition variables[17] each separately explain about half of the size of the quality gaps. Most notable, average ability in the class explains virtually all of ECERS-R gap magnitudes, and classroom composition explains virtually the entire scaffolds gap. Still, even after accounting for these factors, statistically significant gaps of approximately -0.15 to -0.30 *sd* remain on a number of measures.

Thirdly, teacher race and whether a teacher speaks a language other than English in class explain little to none of the magnitude of quality gaps on the CLASS, ECERS-R, and individual time, but explain all of the size of Black-White and Hispanic-White gaps on teachers' beliefs about child-rearing, and in most cases at least half of the size of the free-choice and scaffolding gaps.

After adjusting for all controls, about a third of all estimated gaps remain significant, though mostly only showing coefficients of half to a third as large as the original magnitude. Controlling for state fixed effects explains about 50% of gaps, which suggests that half of quality gap magnitudes result from disadvantaged students being concentrated in states with lower pre-K quality than the average. This may suggest that it is just as important to improve overall quality in some states as it is to ensure equity of access to quality in others. Best seen in the last panel of Figure 4, controlling for all factors and state fixed effects explains virtually all of the magnitude and significance of gaps. This result suggests that most of the inequalities in access may be largely eliminated through policies that (a) ensure that students with limited skills are not

concentrated only in classrooms with each other, and (b) focus on state-level efforts to increase pre-K quality on average.

Question 3: Variation Across States

Results for question three, whether there is between-state variation in quality gap magnitudes, are presented in Figure 4 for the overall CLASS and ECERS-R for the sake of parsimony. Figures for all measures are available upon request. Figures are sorted by the average size of the quality gap across gap groups and by state. It is noteworthy that some state gap estimates are highly imprecisely estimated, although others are quite precise. The imprecision could be explained in part by a small population for a gap group within a state. In a handful of cases, confidence intervals were truncated because they would extend beyond the x-axis range.

*** Insert Figure 4***

There are three main takeaways. First, there is sizable variation in gaps across states on some measures but not others. The most variation is evident on the ECERS-R and CLASS. Further analyses revealed much variation across states in the proportion of the instructional day spent in free choice and scaffolded interactions. Less variation is apparent on measures such as proportion of the day spent as individual time, teachers' belief about child-rearing, and measures of proportion of the day spent on academic content such as science.

Second, these figures can be viewed in clusters of gaps by state within a figure. When a state has a large gap on the measure for one gap group, it tends to have a large gap for all gap groups. This is in part because gaps are necessarily correlated. Black-White and Hispanic-White gaps, for example, both rely on the average classroom quality of White students in the state. Similarly, language status, race, and income are correlated, which results in correlated quality gaps. However, this finding also may be due to the fact that states that tend to have poor quality

programs serving one disadvantaged group also offer similarly low quality programs to other disadvantaged groups.

Third, there are some consistent patterns of state rankings across measures.[18] Across quality gap measures, states six and eleven are among those that tend to display quality gaps favoring poor/minority students, or that are not distinguishable from zero. On the other hand, states eight and two are among those that tend to consistently display large quality gaps favoring non-poor/non-minority students. A robustness check of between-state variation in overall quality indicates that there is much less between-state variation in overall quality than there is in gaps, but that overall quality rankings tend to mimic those of quality gaps.

Question 4: State-Level Factors That Explain Between-State Differences

Finally, this section reports on research question four, about which of three state-level factors—(1) the rate of pre-K expansion since 1995, (2) pre-K spending per child, and (3) residential segregation—correlate with state-level quality gaps. Simple correlational illustrations demonstrate that neither the rate of pre-K expansion nor spending per child is predictive of state-level quality gaps.[19]

However, there are relationships between quality gaps and residential segregation on a number of quality measures, particularly for Black-White and Hispanic-White gaps. Figure 5 displays results of these relationships for the ECERS-R and CLASS. There is a clear relationship between segregation and Black-White and Hispanic-White gaps on the ECERS-R and CLASS; states with the largest gaps favoring White students tend to be those with the highest levels of residential segregation for that gap group, and vice versa. The same relationship is evident for free choice activities. For individual time, the pattern is similar to Black-White gaps. Interestingly, there seems to be a clear positive relationship between Hispanic-White gaps in

teachers' traditional beliefs about child-rearing and residential segregation; suggesting that in areas with higher residential segregation, teachers of more densely populated Hispanic classrooms are more likely to have traditional beliefs about child-rearing. Finally, there is a positive relationship between a state's structural quality gap (particularly the Black-White gap) and the state's residential segregation. There is little or no relationship between structural quality gap and measures of science, letters/sounds, and scaffolds/didactic. Further, there is not enough variation in poor-non-poor residential segregation to strongly detect a relationship.

*** Insert Figure 5***

Discussion

This paper estimates the magnitude of gaps in pre-K quality and classroom process between Black/Hispanic and White, poor and non-poor, and DLL and non-DLL students. It is motivated by the need to understand whether quality pre-K is equitably distributed across students and whether state pre-K has the potential to close the achievement gap. Pre-K can close as much as 50% of the achievement gap if children attend the highest quality programs (Camilli et al., 2010), but the findings in this paper suggest that states will struggle to achieve this level of pre-K success if quality gaps are not closed. Further, the paper has implications for what levers might be able to narrow quality gaps through policy changes. There are four key sets of findings.

Gap Magnitudes

First, results highlight that pre-K quality gaps are large, ranging from about 0.3 to 0.7 *sd*, and mirror the size of achievement gaps at school entry, with the largest gaps observed on the CLASS—a measure of instructional and emotional classroom climate—which is, of all the measures included in this study, the strongest predictor of child outcomes across student groups. One way that pre-K could close achievement gaps is if the pre-K quality experienced by

poor/minority students is higher, on average, than that experienced by non-poor/non-minority students. Findings from this paper suggest that this is not happening, and therefore suggest that state pre-K as it currently exists is unlikely to narrow achievement gaps. It is possible that achievement gaps could still be reduced if disproportionate numbers of poor/minority children attend pre-K and/or if these children have more to gain from pre-K than their more advantaged peers. However, as research demonstrates, while pre-K is generally equally beneficial across student subgroups, it does not tend to be *more* beneficial for poor/minority children than non-poor/non-minority students (Gormley, 2008; Gormley et al., 2005; Magnuson et al., 2006; Weiland & Yoshikawa, 2013). Furthermore, a push for universal over pre-K that is targeted to disadvantaged subgroups, may make it less likely over time that poor and minority children could disproportionately attend pre-K.

There is, however, a caveat to these findings of quality gaps. It is unclear that *all* of these gaps in classroom process are problematic. The magnitude of some gaps may rather be indicative of differences in racial and cultural norms of how children learn. For instance, parenting literature suggests that White parents are more likely to rely on scaffolding approaches to instruction, while Black parents are more likely to engage in didactic and directive interactions (Pellegrini, et al., 1990; Slaughter, 1987; Stipek, 2004). If these are norms that some children are more accustomed to, they may respond more positively than expected to instructional techniques that are not conventionally rated highly by standard measures of quality. Take didactic and directive instruction, for example. There may also be reason to think that disproportionate use of directive instructional approaches with minority students may gain traction on achievement gap closure, as more time is spent directly teaching academic content.

While there is some evidence that directive instruction could actually improve achievement and narrow achievement gaps in the short term (see Chein et al., 2010), it is arguable that such an approach is still unfavorable in the long term. It is likely that activities such as free play stimulate other critical aspects of development, particularly related to outcomes in social skills, creativity, and executive functioning, which matter just as much, if not more, for children's long-term success on outcomes such as school persistence. While short-term achievement gap closure is certainly desirable, equality of long-term life success is ultimately the outcome of interest. Further, while the interpretation of *some* quality gaps as unjust differences in quality is questionable, the persistence of large gaps across measures, even on the most objective measures of quality such as proportion of time spent learning science, and especially on those most strongly predictive of children's development (e.g. the CLASS), indicates that, taken in aggregate, these quality differences are problematic.

What Explains Gaps

Second, results for question two find that structural classroom characteristics explain very little of process quality gap magnitudes, while characteristics of peers in the classroom (both average academic skills and racial/socioeconomic composition) explain 50–65% of gap magnitudes. Teacher race explains the magnitude of some gaps, but not others. There are several key takeaways from these findings. First, the limited explanatory power of structural characteristics is perhaps in part due to little variation across groups within states on these factors, as states tend to regulate structural factors such as teacher degree and wage. Further, as this paper finds, gaps on structural measures are rarely statistically significant. These findings have implications for states' Quality Rating Improvement Systems (QRISs). With increasing use of QRISs among states, the measures that make up these QRISs may be more valid if the

measures are shifted away from those that are not strongly predictive of student outcomes, such as structural measures. Instead, measures that make up the QRISs should shift toward those that link most directly to improved student outcomes, such as the CLASS. This will help ensure that all children have equal access not only to high quality programs, but also to programs that are most likely to yield equity of outcomes across student subgroups. Certainly, the CLASS is far from perfect as a single measure, and more research is needed to ensure that all important classroom inputs that may help narrow achievement and life-long gaps are tracked and measured. For now, the CLASS is the best of existing measures. While a handful of states[20] already incorporate measures like the CLASS into their QRIS, and Head Start is now mandating use of the CLASS, QRIS participation in states is often still low (Tout et al., 2011), and Head Start only represents about 10% of age-eligible 4-year-olds who might attend preschool (ACF, 2015; Barnett et al., 2013). More research is needed to expand the use of best-in-class measures. As the conceptual framework of this paper shows, structural features may be necessary but are not sufficient for ensuring equity of pre-K process quality across student groups. For this reason, they should not be discounted altogether, but they should be weighted less heavily than measures shown to have direct influence on student development. As states begin to tie monetary incentives to QRIS scores, incorporating process measures most strongly linked to improved student outcomes and mandatory participation may be most critical for incentivizing the right quality improvement.

In addition, the finding that average student ability in the classroom explains a substantial portion of gap magnitudes could in part be explained by teachers adapting how and what they teach as a means to target instruction to bring their lower-skilled students up to speed academically. For example, teachers may spend more time teaching letters/sounds if their

students are trailing in literacy skills, and may spend less time in free play for fear that too much unstructured time could slow students' academic progress (e.g. Chein et al., 2010). Another possibility is that process quality measures are sensitive to much more than teachers' behaviors and practices, and rather are also a function of the kinds of students in their classrooms. Poor and minority students are more likely to have behavioral challenges than their non-poor non-minority peers (Reardon & Portilla, 2016), so classrooms with higher proportions of the former group may experience more disruptions. This could lead to less instructional time spent on academic content because teachers are spending more time on classroom management.

Teacher race explain little to none of quality gap magnitudes on the CLASS, ECERS-R, and individual time, but explain the entire Black-White and Hispanic-White gap magnitudes in teachers' beliefs about child-rearing. They also generally explain at least half of free choice and scaffolding gap magnitudes. There are two potential explanations for this. First, this finding is consistent with prior research that Black and Hispanic caregivers are more likely to endorse authoritarian than authoritative beliefs about child-rearing (see Baumrind, 1972; Chao, 2000; Fuller & Clarke, 1994; Howes, 2010; Kermani & Brenner, 2000; Ladson-Billings, 1995). Second, Black and Hispanic teachers may be more likely to teach in classrooms serving higher portions of minority children, as is true in this sample.[21] With minority teachers disproportionately teaching minority students, from the beginning of the school year these teachers may, on average, be more likely to face the challenge of catching their students up academically than are their non-minority teacher counterparts, and may consequently spend less time in free play, worrying that such time could occur at the expense of time spent on academics.

Between-State Variation in Gaps

Third, there is much between-state variation in gap magnitudes, such that some states consistently have large quality gaps across groups, while others have null gaps, or gaps that favor poor/minority preschoolers. This latter finding is somewhat surprising. It may be explained by particularly progressive pre-K policies in some states, such as differential pre-K funding within states for slots serving more disadvantaged children or higher salaries for teachers in higher need areas. It is also possible that in these states, stronger attention is paid to monitoring quality in programs serving the most disadvantaged children. On the other hand, it is possible that states with favorable quality gaps simply have fewer disadvantaged students overall, making it more difficult to experience extreme differential sorting of disadvantaged students into lower quality programs that are more densely populated by other disadvantaged students. This latter explanation is, however, less likely, as marker weighting on figures for question four do not show a strong relationship between minority population size and segregation within states.

State Characteristics and Gaps

Fourth, this paper finds that state-level residential segregation (but not pre-K expansion or spending-per-child) and quality gaps are correlated, particularly in the case of Black-White and Hispanic-White process quality, but not structural quality gaps. These findings suggest that increasing spending per child in pre-K alone is unlikely to do much for gap closure, and that continued pre-K expansion is unlikely to exacerbate the quality gap problem. But, when states are more residentially segregated, Black and Hispanic children tend to experience worse pre-K environments than their White peers. In some ways, this finding is very predictable, as without segregation there necessarily cannot be a quality gap because students of different backgrounds would attend, in expectation, programs of equal quality. Still, this pattern is consistent with what is seen in K–12, where the neighborhood one lives in is tied to the school quality he or she

receives, partly because many effective teachers sort into schools serving lower proportions of Black, poor, low-achieving students (Boyd et al., 2011; Jackson, 2009). These findings suggest that one of two policy approaches might be effective at reducing quality gaps in pre-K and K–12 alike: incentives to recruit higher quality teachers to high-needs neighborhoods, or interventions (e.g. professional development; Pianta et al., 2005) to increase quality among existing teachers.

On the latter point, experimental research shows that early childhood teaching quality is quite malleable. Specifically, a randomized controlled trial of the Head Start REDI program, a professional development intervention involving four days of workshop training and weekly in-class support from a mentor teacher, significantly improved intervention teachers' classroom practices. Intervention teachers talked with children more frequently and in more cognitively complex ways, established a more positive classroom climate, and used more preventive behavior-management strategies, practices that in turn significantly improved children's executive functioning, academic, and social-emotional skills above and beyond their peers in the control group classrooms (Bierman, Domitrovich, Nix, et al., 2008; Bierman, Nix, Greenberg, Blair, & Domitrovich, 2008; Domitrovich, Gest, Gill, Bierman, Welsh, & Jones, 2009). Similar improvements in the quality of teacher-child interactions were found in the randomized study of MyTeachingPartner—a web-based system of professional development resources, including video exemplars and web-medicated consultation on specific dimensions of interactions (Pianta, Mashburn, Downer, Hamre, & Justice, 2008). Taken together, these studies demonstrate that the malleability of quality is more heavily influenced by how workforce development resources are deployed than by teacher selection factors alone.

Limitations and Future Research

Despite the study contributions, there remain limitations worth noting. First, findings about classroom factors that explain quality gap magnitudes (and policy implications associated with these findings) should be interpreted with caution, as they are not causally estimated. It is possible that other unobservable school- or classroom-level factors explain quality gap magnitudes.

Second, the current data set is close to 15 years old, which calls into question whether quality gaps are still as large as they were at the time of data collection. However, while the average quality preschool programs may have improved since 2004, this paper shows that most disadvantaged children attended public preschools that were *far inferior* to those available to their more advantaged peers. It is unlikely that these quality gaps have been fully remediated since the time of data collection. If they had been, we would expect to have observed at least some narrowing of achievement gaps over that time frame as well. We have seen only modest narrowing of achievement gaps, at best, and this narrowing could be attributable to a number of factors (see Reardon & Portilla, 2016).

Third, while compelling, the between-state variation findings and relationship between segregation and gaps warrant further research. Although this study represents the states enrolling the majority of children in public pre-K, data is only available for 11 states, and therefore cannot be extrapolated to other states or to the entire United States. The small sample of states made it difficult to quantitatively investigate the relationship between state-level factors and quality gaps. For this reason, future efforts to explore such relationships with a larger sample of states is warranted. Still, this evidence is suggestive that segregation may be a driver of disparities in pre-K quality across groups.

Finally, while the number and magnitude of gaps are certainly indicative of unjust and inequitable experiences for poor and minority children on the whole, more research is necessary to determine which of these indicators are interpretively the most problematic. More specifically, research is necessary to parse out the benefits and drawbacks of indicators of quality such as free play and scaffolded instruction versus individual time and didactic instruction across different student groups.

Notes

1. Dual language learners (DLL) are students whose home language is not English. DLL preferred because children are not yet classified as English Language Learners in preK.
2. Six of the 11 state pre-K programs included in the current study.
3. Although new state pre-K programs have been created, and states have expanded pre-K programs since these data were collected, these 11 states still represent close to 60% of all children enrolled in and funds allocated to pre-K (Barnett et al., 2013).
4. A full list of items and descriptors for available upon request.
5. Kappa is a reliability statistic that captures inter-rater agreement. It is considered a more conservative measure of agreement, as it takes into account the probability of reaching agreement by chance.
6. Figures were inflation adjusted to 2014 dollars.
7. This variable was the combination of two variables, whether the child was in preschool, and whether it was public or private.
8. Sampling weights took the form of the inverse probability of selection to ensure unbiased parameter estimation.
9. Jackknife standard errors should produce results asymptotically the same as bootstrap standard errors, and are ideal to estimate standard errors with complex survey designs.
10. Models were equivalent in form to those described in reference to question two below, except without the covariates X_c and state fixed effects Γ_s.
11. Subpopulation estimation involves computing point and variance estimates for part of the population (e.g. just between Black and White students), but is different from just restricting the sample to the observations within the subpopulation prior to running the model because variance estimation for survey data measures sample-to-sample variability (West, Berglund, & Heeringa, 2008).
12. Signs on gaps presented in tables and graphs represent the direction indicated by the order of subgroup names. For example, in the case of Black-White, gaps are calculated as Black relative to White. For this reason most negative gaps indicate that Black students experience lower quality pre-K, on average, than their White peers.
13. Reference Table 2 for a full list of the proportion of the day spent on activities/in settings.
14. Note that this gap is positive because class-size is an increasing number. But larger class sizes are generally viewed unfavorably, so still, this gap favors White students.
15. Includes class size, teacher-child ratio, teacher degree, years of experience, whether class is full-day, hours/week spent in class, and teacher wage.
16. As measured by average student vocabulary, math, and competence in the classroom.
17. Proportion of students in the class that are Black, Hispanic, DLL, or poor.
18. The sorting of states may not be consistent across all quality measures because some states are better performers on some measures than others, but also because quality gaps for any two states ranked next to each other are rarely statistically distinguishable from each other, so some of what determines where they rank is due to random noise.
19. A table with the rates of pre-K expansion by state and state pre-K spending by state available upon request.
20. Including California, Arizona, Oklahoma, Virginia, and Georgia.
21. On average, 78% of students in a Black teacher's class are Black/Hispanic and 70% are poor, compared to 5% and 42%, respectively in the case of White teachers.

References

Administration for Children and Families (ACF) (2015). *Head Start program facts fiscal year 2015*. Retrieved from https://eclkc.ohs.acf.hhs.gov/hslc/data/factsheets/docs/head-start-fact-sheet-fy-2015.pdf

Barnett, W.S. (2011). Effectiveness of early educational intervention. *Science, 333*, 975–978.

Barnett, W. S., Carolan, M., & Johns, D. (2013). *Equity and excellence: African-American children's access to quality preschool.* Center for Enhancing Early Learning Outcomes & National Institute for Early Education Research.

Barnett, W. S., Carolan, M. E., Squires, J. H., & Brown, K. C. (2013). *The state of preschool 2013: State preschool yearbook.* New Brunswick, NJ: National Institute for Early Education Research.

Barnett, W.S., Robin, K. B., Hustedt, J. T., & Shulman, K.L. (2003). *The state of preschool 2003: State preschool yearbook.* New Brunswick, NJ: National Institute for Early Education Research.

Bassok, D., Fitzpatrick, M., Greenberg, E., & Loeb, S. (2013). *The extent of within- and between-sector quality differences in early childhood education and care.* Unpublished manuscript.

Bassok, D., & Galdo, E. (2015). *Inequality in preschool quality? Community-level disparities in access to high quality learning environments.* Unpublished manuscript.

Baumrind, D. (1972). An exploratory study of socialization effects on black children: Some Black-White comparisons. *Child Development, 43*, 261–267.

Bierman, K. L., Domitrovich, C. E., Nix, R. L., Gest, S. D., Welsh, J. A., Greenberg, M. T., Blair, C., Nelson, K. E., & Gill, S. (2008). Promoting academic and social-emotional school readiness: The Head Start REDI program. *Child Development, 79*, 1802–1817.

Bierman, K. L., Nix, R. L., Greenberg, M. T., Blair, C., & Domitrovich, C. E. (2008). Executive functions and school readiness intervention: Impact, moderation, and medication in the Head Start REDI program. *Development and Psychopathology, 20*, 821–843.

Boyd, D., Lankford, H., Loeb, S., Ronfeldt, M., & Wyckoff, J. (2011). The effect of school neighborhoods on teachers' career decisions. In G. J. Duncan & R. J. Murnane (Eds.), *Whither opportunity* (pp. 377–395). New York, NY: Russell Sage.

Burchinal M. R., Howes, C., Pianta, R., & Barbarin, O. A. (2008). Predicting child outcomes at the end of kindergarten from the quality of pre-kindergarten teacher-child interactions and instruction. *Applied Developmental Science, 12*, 140–153.

Burchinal, M., Vandergrift, N., Pianta, R., & Mashburn, A. (2010). Threshold analysis of association between child care quality and child outcomes for low-income children in pre-kindergarten programs. *Early Childhood Research Quarterly, 25,* 166–176.

Camilli, G., Vargas, S., Ryan, S., & Barnett, W. S. (2010). Meta-analysis of the effects of early education interventions on cognitive and social development. *Teachers College Record, 112*, 579–620.

Campbell, F. A., Ramey, C. T., Pungello, E., Sparling, J., & Miller-Johnson, S. (2002). Early childhood education: Young adult outcomes from the Abecedarian Project. *Applied Developmental Science, 6*, 42–57.

Chao, R. K. (2000). The parenting of immigrant Chinese and European American mothers: Relations between parenting styles, socialization goals, and parenting practices. *Journal of Applied Developmental Psychology, 21*, 233–248.

Chien, N. C., Howes, C., Pianta, R. C., Burchinal, M., Ritchie, S., Bryant, D. M., Clifford, R. M., Early, D. M., & Barbarin, O. A. (2010). Children's classroom engagement and school readiness gains in prekindergarten. *Child Development, 81*, 1534–1549.

Clarke-Stewart, K. A., Lowe Vandell, D., Burchinal, M., O'Brien, M., & McCartney, K. (2002). Do regulable features of child-care homes affect children's development? *Early Childhood Research Quarterly, 17*, 52–86.

Clifford, R. M., & Reszka, S. S. (2010). *Reliability and validity of the Early Childhood Environment Rating Scale.* Chapel Hill, NC: Frank Porter Graham Child Development Institute.

Domitrovich, C. E., Gest, S. D., Gill, S., Bierman, K. L., Welsh, J. A., Jones, D. (2009). Fostering high-quality teaching with enriched curriculum and professional development support: The Head Start REDI program. *American Educational Research Journal, 46*, 567–597.

Duncan, G. J., & Magnuson, K. A. (2005). Can family socioeconomic resources account for racial and ethnic test score gaps? *Future of Children, 15*, 35–54.

Dunn, L. M., & Dunn, D. M. (1997). *Peabody picture vocabulary test* (3rd ed.). Circle Pines, MN: American Guidance Service, Inc.

Early, D. M., Iruka, I. U., Ritchie, S., Barbarin, O. A., Winn, D. C., Crawford, G. M., . . . Pianta, R. C. (2010). How do pre-kindergarteners spend their time? Gender, ethnicity, and income as predictors of experiences in pre-kindergarten classrooms. *Early Childhood Research Quarterly, 25*, 177–193.

Early, D. M., Maxwell, K. L., Burchinal, M., Alva S., Bender R. H., Bryant D., . . . Zill, N. (2007). Teachers' education, classroom quality, and young children's academic skills: Results from seven studies of preschool programs. *Child Development, 78*, 558–580.

Fuller, B. & Clarke, P. (1994). Raising school effects while ignoring culture? Local conditions and the influence of classroom tools, rules, and pedagogy. *Review of Educational Research, 64,* 119–157.

Garces, E., Thomas, D., & Currie, J. (2002). Longer term effects of Head Start. *The American Economic Review, 92*, 69–123.

Gormley, W.T. (2008). The effects of Oklahoma's pre-K program on Hispanic children. *Social Science Quarterly, 89*, 916-936.

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

Gould, W. (1999). *What is the effect of specifying aweights with regress?* (Stata Technical Bulletin 20). Retrieved from http://www.stata.com/support/faqs/statistics/analytical-weights-with-linear-regression/

Harms, T., Clifford, R. M., & Cryer, D. (2005). *Early childhood environmental rating scale revised edition*. New York: Teachers College Press.

Hightower, A. D., Work, W. C., Cowen, E. L., Lotyczewski, B. S., Spinell, A. P., Guare, J. C., & Rohrbeck, C.A. (1986). The Teacher-Child Rating Scale: A brief objective measure of elementary children's school problem behaviors and competencies. *School Psychology Review, 15*, 393–409.

Howes, C. (2010). *Culture and child development in early childhood programs: Practices for quality education and care*. New York, NY: Teachers College Press.

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

Hustedt, J. T., & Barnett, S. (2011). Financing early childhood education programs: State, federal, and local issues. *Educational Policy, 25*, 167–192.

Jackson, C. K. (2009). Student demographics, teacher sorting, and teacher quality: Evidence from the end of school desegregation. *Journal of Labor Economics, 27*, 213–256.

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

Kermani, H. & Brenner, M. E. (2000). Maternal scaffolding in the child's zone of proximal development across tasks: Cross-cultural perspectives. *Journal of Research in Childhood Education, 15*, 30–52.

Kolenikov, S. (2010). Resampling variance estimation for complex survey data. *The Stata Journal, 10*, 165–199.

Ladson-Billings, G. (1995). But that's just good teaching! The case for culturally relevant pedagogy. *Theory into Practice, 34*, 159–165.

Lankford, H. Loeb, S., & Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: A descriptive analysis. *Educational Evaluation and Policy Analysis, 24*, 37–62.

Leak, J., Duncan, G., Li, W., Magnuson, K., Schindler, H., & Yoshikawa, H. (2010, November). *Is timing everything? How early childhood education program cognitive and achievement impacts vary by starting age, program duration and time since the end of the program.* Paper presented at the meeting of the Association for Policy Analysis and Management, Baltimore, MD.

Lee, V. E., & Burkam, D. T. (2002). *Inequality at the starting gate.* Washington, DC: Economic Policy Institute.

LoCasale-Crouch, J., Konold, T., Pianta, R., Howes, C., Burchinal, M. Bryant, D., . . . Barbarin, O. (2007). Observed classroom quality profiles in state-funded pre-kindergarten programs and associations with teacher, program, and classroom characteristics. *Early Childhood Research Quarterly, 22*, 3–17.

Loeb, S., & Bassok, D. (2008). Early childhood and the achievement gap. In H. F. Ladd & E. B. Fiske (Eds.), *Handbook of research in education finance and policy* (pp. 497–516). New York, NY: Routledge.

Magnuson, K. A., Ruhm, C., & Waldfogel, J. (2007). The persistence of preschool effects: Do subsequent classroom experiences matter? *Early Childhood Research Quarterly, 22*, 18-38.

Magnuson, K., Lahaie, C., & Waldfogel, J. (2006). Preschool and school readiness of children of immigrants. *Social Science Quarterly, 87,* 1241-1262.

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

Morrison, F. J., & Connor, C. M. (2002). Understanding schooling effects on early literacy: A working research strategy. *Journal of School Psychology, 40*, 493–500.

NICHD Early Child Care Research Network. (2002). Child-care structure → process → outcome: Direct and indirect effects of child-care quality on young children's development. *Psychological Science, 13*, 199–206.

Nores, M. & Barnett, W. S. (2014). *Preschool quality in structure and process. Why the same story can have different endings.* Unpublished manuscript.

Peisner-Feinberg, E.S., Burchinal, M.R., Clifford, R.M., Culkin, M.L., Howes, C., Kagan, S., & Yazejian, N. (2001). The relation of preschool child-care quality to children's cognitive and social developmental trajectories through second grade. *Child Development, 72*, 1534–1553.

Pellegrini, A., Perlmutter, J., Galda, L., & Brody, G. (1990). Joint reading between Black Head Start children and their mothers. *Child Development, 61*, 443–453.

Pianta, R. C. (2006). Teacher-child relationships and early literacy. In D. Dickinson & S. Newman (Eds.), *Handbook of early literacy research* (Vol. 2, pp. 149–162). New York: The Guilford Press.

Pianta, R. C., & Howes, C. (Eds.) (2009). *The promise of pre-K*. Baltimore, MD: Brookes Publishing.

Pianta, R. C., Howes, C., Burchinal, M., Bryant, D., Clifford, R., Early, D., & Barbarin, O. (2005). Features of pre-kindergarten programs, classrooms, and teachers: Do they predict observed classroom quality and child-teacher interactions? *Applied Developmental Science, 9*, 144–159.

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System Pre-K*. Baltimore, MD: Brookes Publishing.

Pianta, R. C., Mashburn, A. J., Downer, J. T., Hamre, B. K., & Justice, L. (2008). Effects of web-mediated professional development resources on teacher-child interactions in pre-kindergarten classrooms. *Early Childhood Research Quarterly, 2008*, 431–451.

Pianta, R. C., Steinberg, M. S., & Rollins, K. B. (1995). The first two years of school: Teacher-child relationships and deflections in children's classroom adjustment. *Development and Psychopathology, 7*, 295–312.

Reardon, S. F., & Galindo, C. (2009). The Hispanic-White achievement gap in math and reading in the elementary grades. *American Educational Research Journal, 46*, 853–891.

Reardon, S. F., & Portilla, X. A. (2016). Recent trends in income, racial, and ethnic school readiness gaps at school entry. *AERA Open, 2*, 1–18.

Reardon, S. F., & Robinson, J. P. (2008). Patterns and trends in racial/ethnic and socioeconomic academic achievement gaps. In H. F. Ladd & E. B. Fiske (Eds.), *Handbook of research in education finance and policy* (pp. 497–516). New York, NY: Routledge.

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

Sabol, T. J., & Pianta, R. C. (2014). Do standard measures of preschool quality used in statewide policy predict school readiness? *Education Finance and Policy, 9,* 116-164.

Schaefer, E., & Edgerton, M. (1985). Parental and child correlates parental modernity. In I. E. Sigel (Ed.), *Parental belief systems* (pp. 121–147). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Schweinhart, L. J., Montie, J., Xiang, Z., Barnett, W. S., Belfield, C. R., & Nores, M. (2005). *Lifetime effects: The High/Scope Perry Preschool study through age 40*. Ypsilanti: High/Scope Press.

Slaughter, D. (1987). The home environment and academic achievement of Black American children and youth: An overview. *Journal of Negro Education, 56*, 3–20.

Stipek, D. (2004). Teaching practices in kindergarten and first grade: Different strokes for different folks. *Early Childhood Research Quarterly, 19*, 548–568.

Tout, K., Starr, R., Soli, M., Moodie, S., Kirby, G., & Boller, K. (2011). *Compendium of quality rating systems and evaluations.* Prepared for the Office of Planning, Research, & Evaluation, Washington, D.C. Retrieved from http://www.acf.hhs.gov/sites/default/files/opre/qrs_compendium_final.pdf

Weiland, C., & Yoshikawa, H. (2013). Impacts of a prekindergarten program on children's mathematics, language, literacy, executive function, and emotional skills. *Child Development, 84*, 2112–2130.

West, B. T., Berglund, P., & Heeringa, S. G. (2008). A closer examination of subpopulation analysis of complex-sample survey data. *The Stata Journal, 8*, 520–531.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III: Tests of achievement*. Itasca, IL: Riverside Publishing.

Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M. R., Espinosa, L. M., Gormley, W.T., . . . Zaslow, M. J. (2013). *Investing in our future: The evidence base on preschool education.* Foundation for Child Development. Retrieved from https://www.fcd-us.org/the-evidence-base-on-preschool

Tables

Table 1. Relative predictive validity of measures of classroom quality

Measure	Subscale or Factor	Type of measure	Significant direct predictor of child outcomes?[1]	For more information see
ECERS	Language & Interactions	Process	Med (Academic)	Peisner-Feinberg et al., 2001; Burchinal et al., 2008; Clifford & Reszka, 2010
ECERS	Provisions for Learning	Structure (E.g. materials)	Low	Sabol & Pianta, 2014; Mashburn et al., 2008
CLASS	Emotional Climate	Process	High (Social-Emotional)	Mashburn et al., 2008
CLASS	Instructional Climate	Process	High (Academic)	Mashburn et al., 2008
Snapshot	Activity (e.g. Science)	Content	Med	Howes et al., 2008
Snapshot	T-C Interactions (e.g. Scaffolded)	Process	Med	Chien et al., 2010
Structure	T degree, years of experience, etc	Structure	Low	Early et al., 2007

[1] Some evidence that measures that are low direct predictors of child outcomes may be significant indirect predictors of child outcomes through their influence on classroom process (e.g. see Nores & Barnett, 2014)

Table 2. Descriptive statistics: Means and standard deviations of unstandardized quality measures, child, and classroom characteristics

Classroom Quality Measures		Classroom Explanatory Variables: Average student characteristics (Question 2)		Classroom Explanatory Variables: Average teacher and structural characteristics (Question 2)	
	mean/sd		mean/sd		mean/sd
CLASS (Total Score)	4.464	Proportion black	0.192	Ratio (Children enrolled to staff)	8.846
	(0.622)		(0.300)		(3.218)
CLASS (Emotional Climate)	5.512	Proportion white	0.402	Class Size	17.477
	(0.678)		(0.367)		(3.913)
CLASS (Instructional Climate)	2.036	Proportion latino	0.27	Teacher Years of Education	16.054
	(0.736)		(0.350)		(1.780)
ECERS (Total Score)	3.828	Proportion poor	0.541	Teacher Yrs Experience in pre-K	8.552
	(0.793)		(0.316)		(6.782)
ECERS (Language/Interactions)	4.659	Proprtion DLL	0.205	Teacher Degree ECE specific	0.576
	(1.161)		(0.318)		(0.495)
ECERS (Provisions for Learning)	3.709	Average maternal education	12.855	Full-Day (vs. Half-Day)	0.531
	(0.964)		(1.409)		(0.499)
Free Choice	0.293	Average vocabulary prescore (PPVT/TVIP)	91.936	Hours Per Week in Class	24.496
	(0.156)		(11.033)		(12.680)
Individual Time	0.044	Average applied problems prescore (WJ)	95.554	Teacher Hourly Wage (2014 $)	26.423
	(0.065)		(10.728)		(12.600)
Didactic	0.291	Average social competence (TCRS)	3.439	Teacher is Black	0.138
	(0.123)		(0.518)		(0.345)
Scaffolded	0.085			Teacher is Asian	0.042
	(0.064)				(0.200)
Science	0.071			Teacher is Latino/a	0.166
	(0.056)				(0.372)
Letters/Sounds	0.039			Teacher is White	0.696
	(0.041)				(0.461)
Math	0.07			Instruction in Other Lang	0.345
	(0.045)				(0.476)
Traditional Child-Rearing Views	38.281				
	(9.356)				
N (Classrooms)	647		647		647

Table 3. Classroom quality gaps between groups

Quality Indicator		Black-White	Hispanic-White	Poor-Non-poor	DLL-Non-DLL
		Classroom climate (CLASS)			
Classroom climate (CLASS)	Total Score	-.659***	-.374***	-.377***	0.02
		[.164]	[.118]	[.096]	[.094]
	Emotional Climate Factor	-.662***	-.238*	-.296***	.144+
		[.193]	[.118]	[.108]	[.077]
	Instructional Climate Factor	-.334***	-.313***	-.298***	-0.092
		[.117]	[.099]	[.065]	[.078]
		Classroom environment (ECERS)			
Classroom environment (ECERS)	Total Score	-.567***	-.523***	-.43***	-.353***
		[.157]	[.143]	[.095]	[.102]
	Language/Interactions Factor	-.647***	-.471***	-.428***	-.17*
		[.149]	[.142]	[.1]	[.085]
	Provisions for Learning Factor	-.421**	-.547***	-.395***	-.572***
		[.17]	[.162]	[.102]	[.101]
		Proportion of classroom time (Snapshot)			
Proportion of classroom time (Snapshot)	Free Choice	-.501***	-.582***	-.396***	-.522***
		[.142]	[.147]	[.092]	[.13]
	Individual Time	.394***	.348*	.359***	.423+
		[.136]	[.162]	[.067]	[.22]
	Didactic	.25+	.495***	.212***	.527***
		[.132]	[.129]	[.081]	[.175]
	Scaffolded	-.301*	-.111	-.199*	0.05
		[.13]	[.101]	[.099]	[.099]
	Science	-.469***	-.302**	-.259***	-0.218
		[.075]	[.118]	[.066]	[.152]
	Letters/Sounds	.306***	.394***	.182*	.424+
		[.104]	[.12]	[.082]	[.227]
	Math	-0.093	0.037	-0.059	.252**
		[.09]	[.082]	[.054]	[.101]
		Structural factors			
Structural factors	Class Size (Reverse Coded)	.229+	.4***	0.093	0.144
		[.118]	[.122]	[.065]	[.114]
	Years Experience in Pre-K	0.024	-.243*	-.138*	-.396***
		[.166]	[.114]	[.069]	[.089]
	Ratio (Children enrolled to staff)	.259+	0.312	.157+	0.248
		[.154]	[.193]	[.089]	[.209]
	Teacher Years of Education	-0.003	0.093	-0.02	.219*
		[.162]	[.146]	[.086]	[.105]
		Teachers' beliefs			
Teacher beliefs	Traditional Child-Rearing Views	.481***	.367***	.186+	.288***
		[.152]	[.104]	[.097]	[.081]
	N (Students)	12,334	12,334	12,334	12,334

Figures

Figure 1. Conceptual Framework: How state policy and classroom structure, process, and other factors influence child outcomes

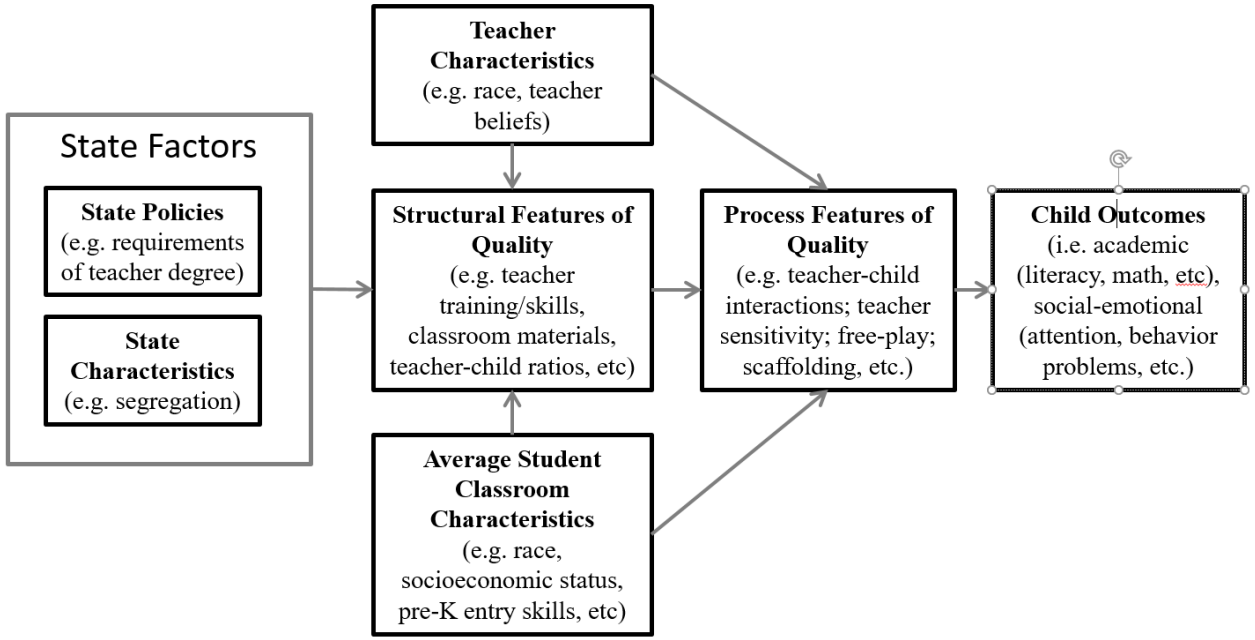
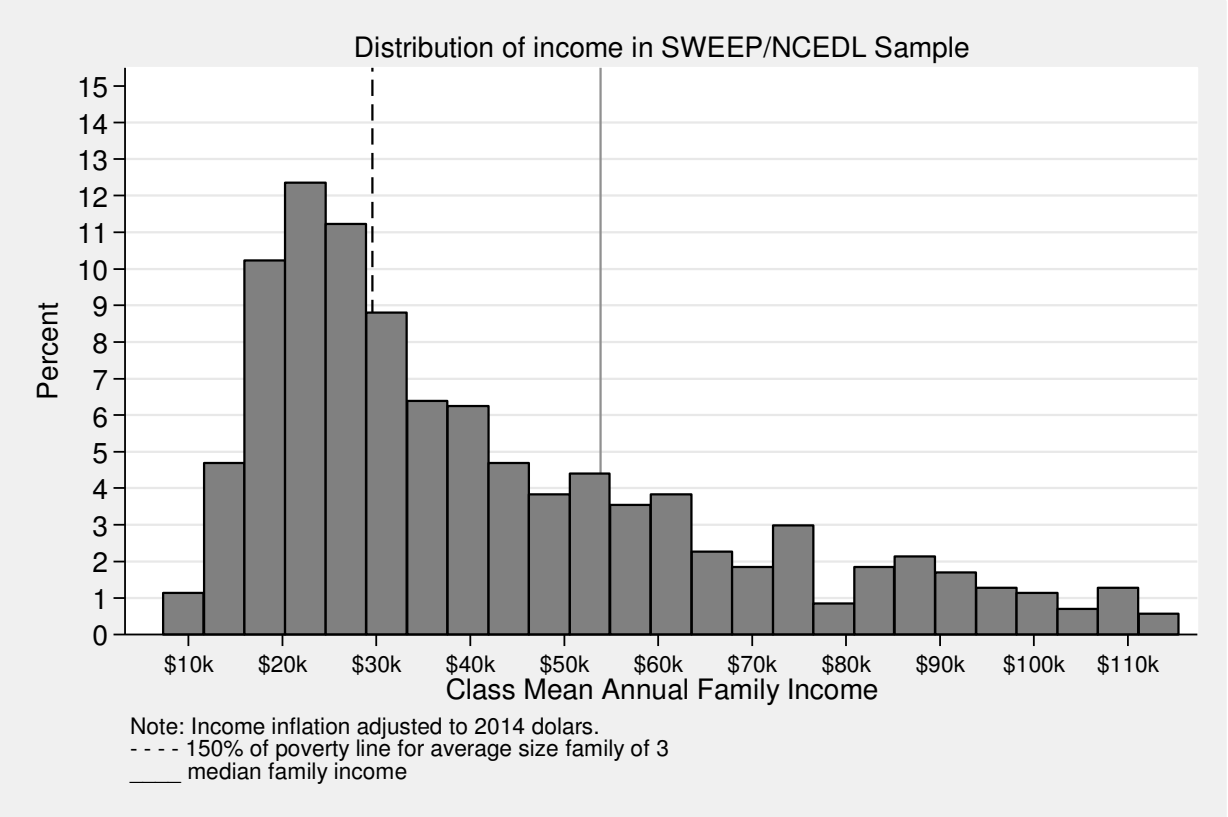
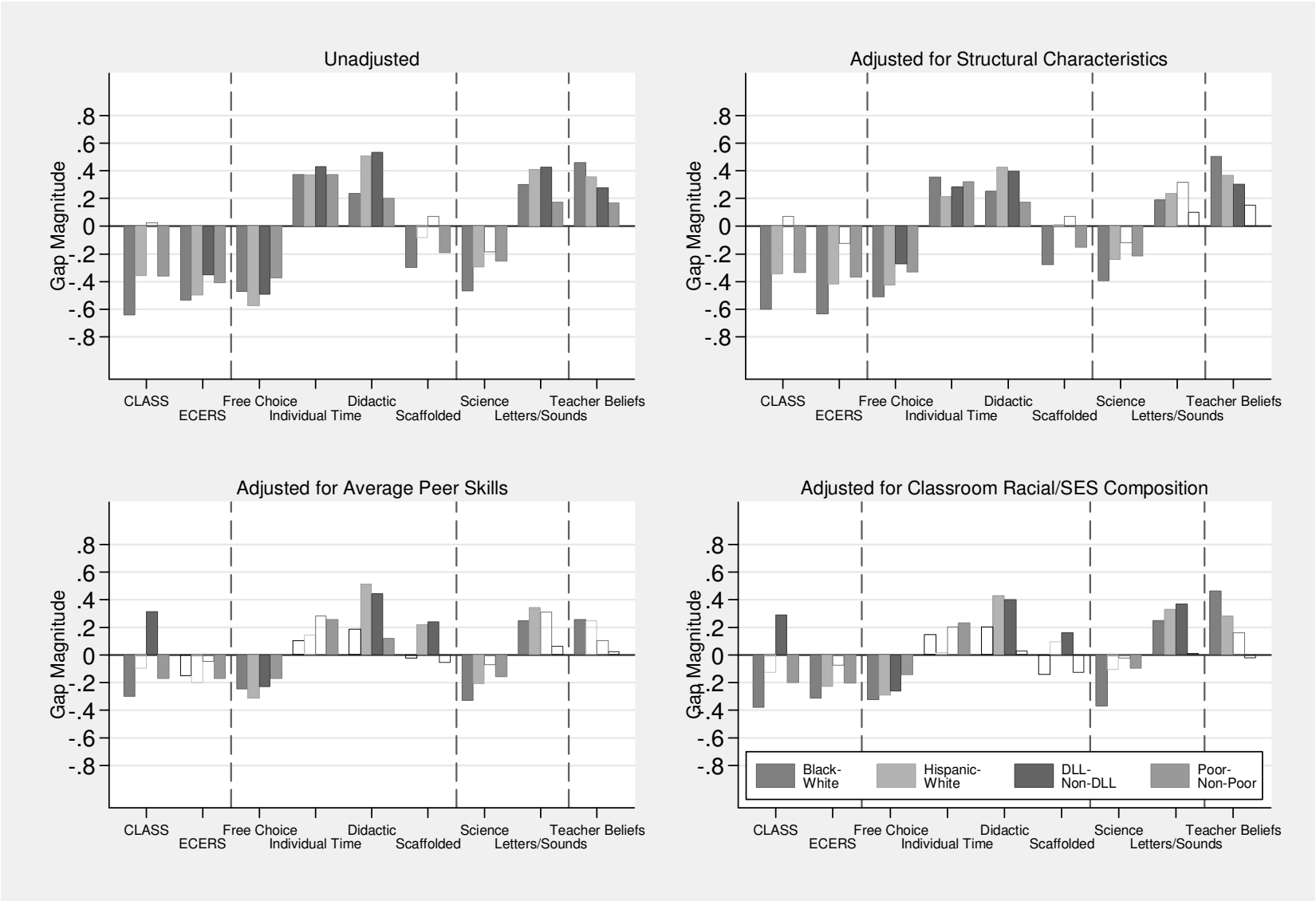


Figure 2. Income Distribution of the Sample



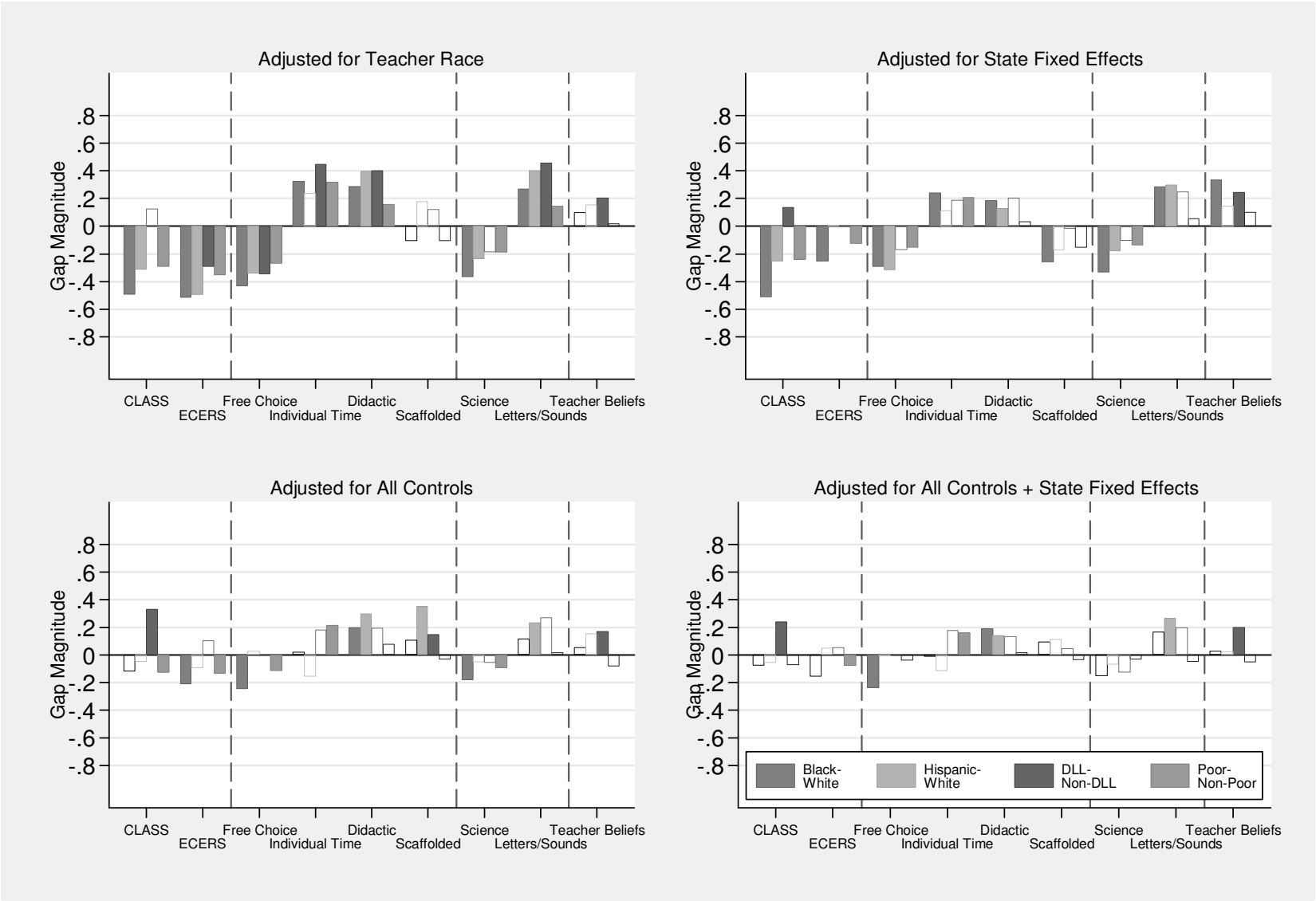
Note: Income inflation adjusted to 2014 dollars
----- 150% of poverty line for average family size of 3; ____ median family income

Figure 3a. Quality gaps, unadjusted and adjusted for classroom-level factors



*Note: Solid bars represent a significant quality gap. Hollow bars represent a non-significant gap.

Figure 3b. Quality gaps, unadjusted and adjusted for classroom-level factors continued



*Note: Solid bars represent a significant quality gap. Hollow bars represent a non-significant gap.

Figure 4. Between state variation in overall ECERS & CLASS scores

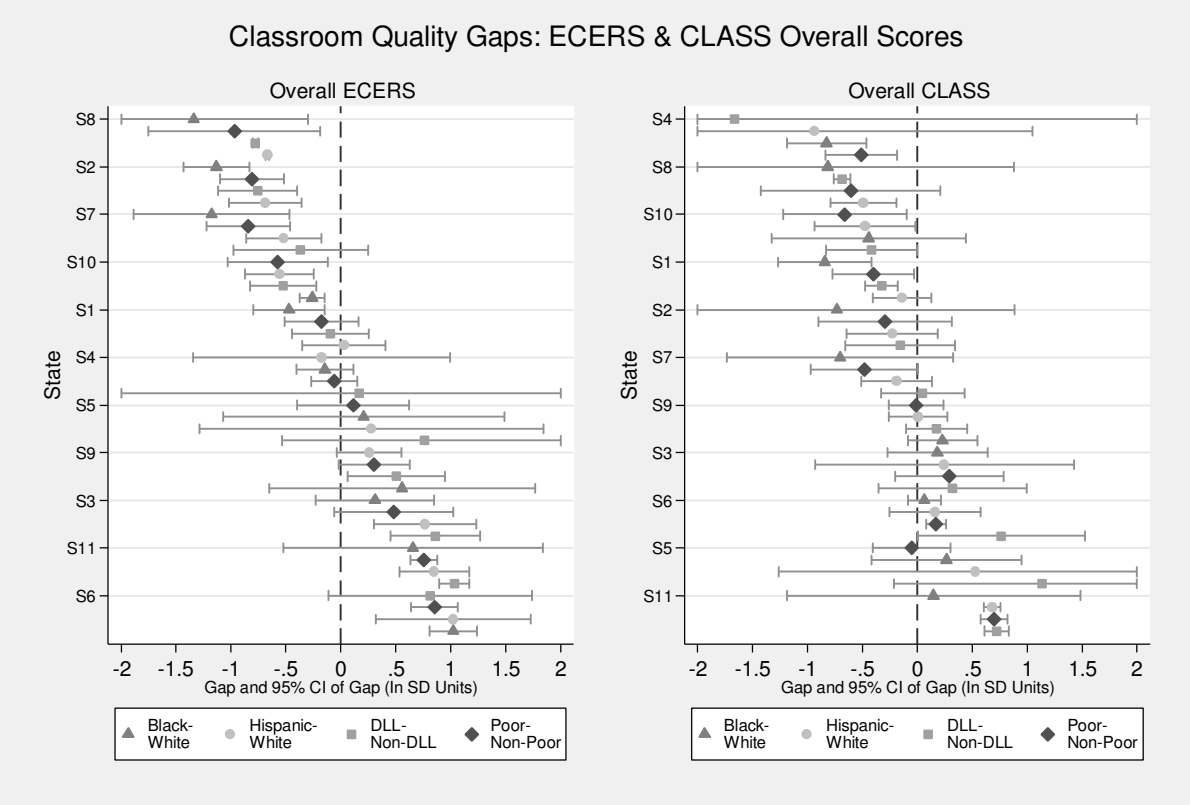
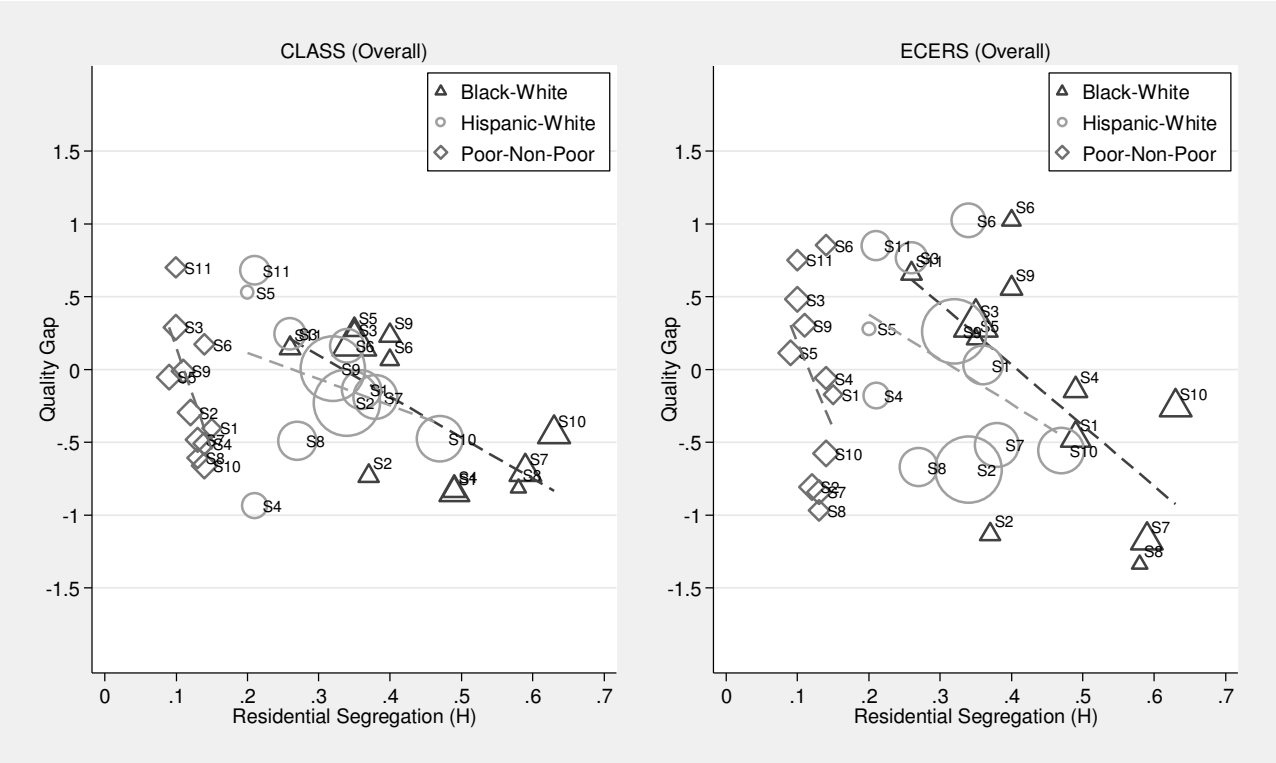


Figure 5. Quality gaps by level of residential segregation and state



*Size of shapes represent the size of the Black, Hispanic, or poor population in that state