

# Research BRIEF

Nikki Aikens, Charles Bush, Philip Gleason, Lizabeth Malone, and Louisa Tarullo

## Tracking Quality in Head Start Classrooms: FACES 2006 to FACES 2014

### INTRODUCTION

This brief examines data from the Head Start Family and Child Experiences Survey (FACES) for trends in observed classroom quality and selected teacher characteristics<sup>1</sup> (such as credentials and professional development activities) between 2006 and 2014.<sup>2</sup> We also examine whether changes in selected teacher characteristics are related to changes in classroom quality.

We examine four key questions to understand change in observed classroom quality and teacher characteristics:

1. Has the observed quality of Head Start classrooms changed from FACES 2006 to FACES 2014, or from FACES 2009 to FACES 2014?
2. Are there changes in mentoring in Head Start programs from FACES 2006 to 2014? Specifically, how many teachers report that they have a mentor? Who provides mentoring?
3. Have the levels of education of Head Start teachers changed from FACES 2006 to 2014?
4. Are changes in observed classroom quality across cohorts explained by selected teacher characteristics (prevalence of mentoring, who provides mentoring, and level of education)?

The findings that follow do not indicate causation. In other words, we cannot attribute trends in classroom quality to specific policy and practice initiatives undertaken by OHS. Similarly, we are not able to conclude that selected teacher characteristics cause trends in observed classroom quality. Rather, the analyses identify whether selected teacher characteristics help to explain trends in observed quality by examining whether trends occurring across the same periods of time in observed quality and teacher characteristics are correlated with one another. Findings may help to inform future work on factors that may contribute to improvements in classroom quality. In Box 1 we describe the FACES design and classroom quality measures used in the current analyses. Box 2 describes the approach to the analyses.

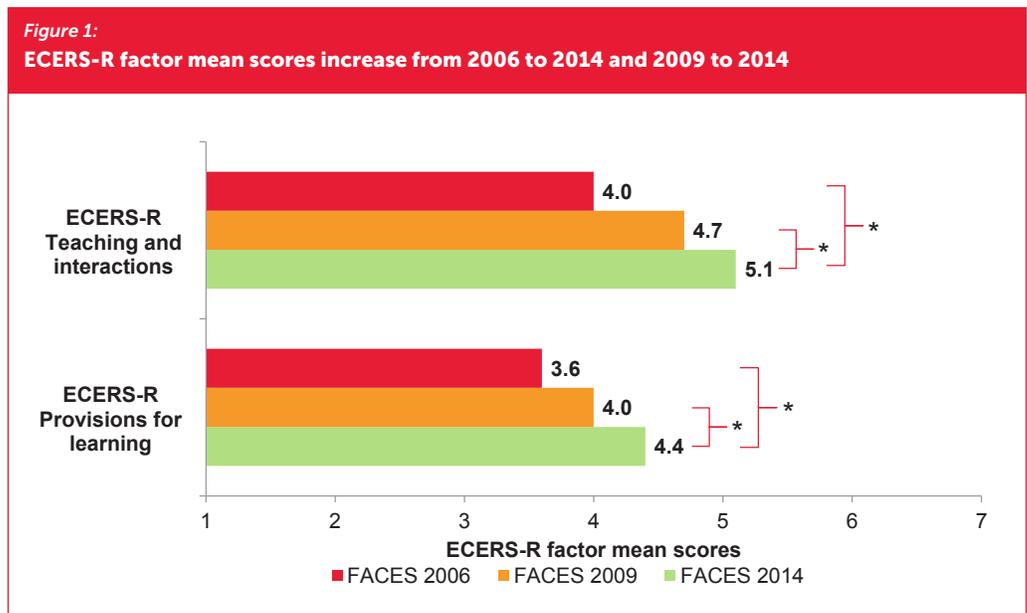
**KEY FINDINGS**

**Has the observed quality of Head Start classrooms changed from FACES 2006 to FACES 2014, or from FACES 2009 to FACES 2014?**

Average scores on the two factors of the Early Childhood Environment Rating Scale–Revised (ECERS-R, Harms et al. 1998), Provisions for Learning and Teaching and Interactions, improve across cohorts, including from FACES 2006 to 2014 and from FACES 2009 to 2014 (Figure 1).

Similarly, based on publisher-developed categories, there are improvements in ECERS-R Provisions for Learning and Teaching and Interactions factor scores from FACES 2006 to 2014, with fewer classrooms scoring in the inadequate and minimal ranges and more scoring in the good or excellent range over time (Figures 2 and 3).

From FACES 2009 to 2014 on the ECERS-R Teaching and Interactions factor, fewer classrooms score in the inadequate range. There were no statistically significant changes in the other two categories of quality, however. Also from 2009 to 2014 on the Provisions for Learning factor, fewer classrooms score in the minimal range, and more classrooms score in the good/excellent range.

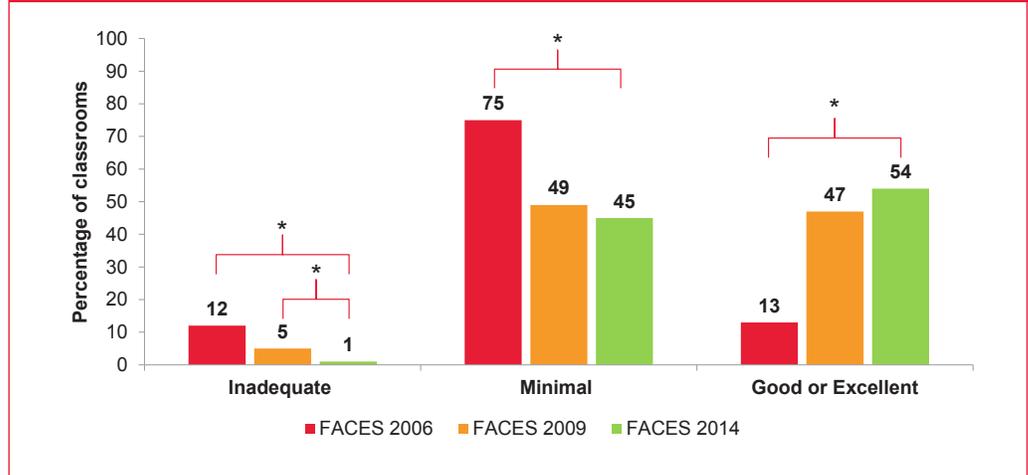


Source: Spring 2007, 2010, and 2015 FACES Classroom Observation.

Notes: Statistics are weighted to represent all Head Start classrooms in the cohort year. Data are drawn from Table B.1 in the accompanying technical report (Aikens et al. 2016).

\* Asterisk indicates that the change over time is statistically significant at the  $p \leq .05$  level.

**Figure 2:**  
**ECERS-R Teaching and Interactions: Classrooms in good/excellent range increase from 2006 to 2014**



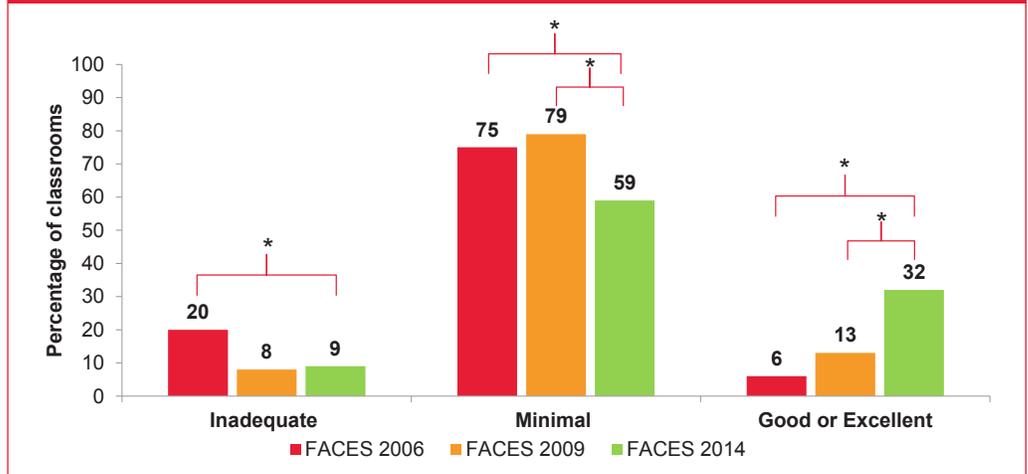
Source: Spring 2007, 2010, and 2015 FACES Classroom Observation.

Notes: Statistics are weighted to represent all Head Start classrooms in the cohort year.

Scores on the ECERS-R factors range from 1 to 7 with publisher-developed categories including inadequate (scores equal to 1 or 2), minimal (scores equal to 3 or 4), good (scores equal to 5 or 6), and excellent (scores equal to 7) quality. Data are drawn from Table B.1a in the accompanying technical report (Aikens et al. 2016).

\* Asterisk indicates that the change over time is statistically significant at the  $p \leq .05$  level.

**Figure 3:**  
**ECERS-R Provisions for Learning: Classrooms in good/excellent range increase from 2006 to 2014 and 2009 to 2014**



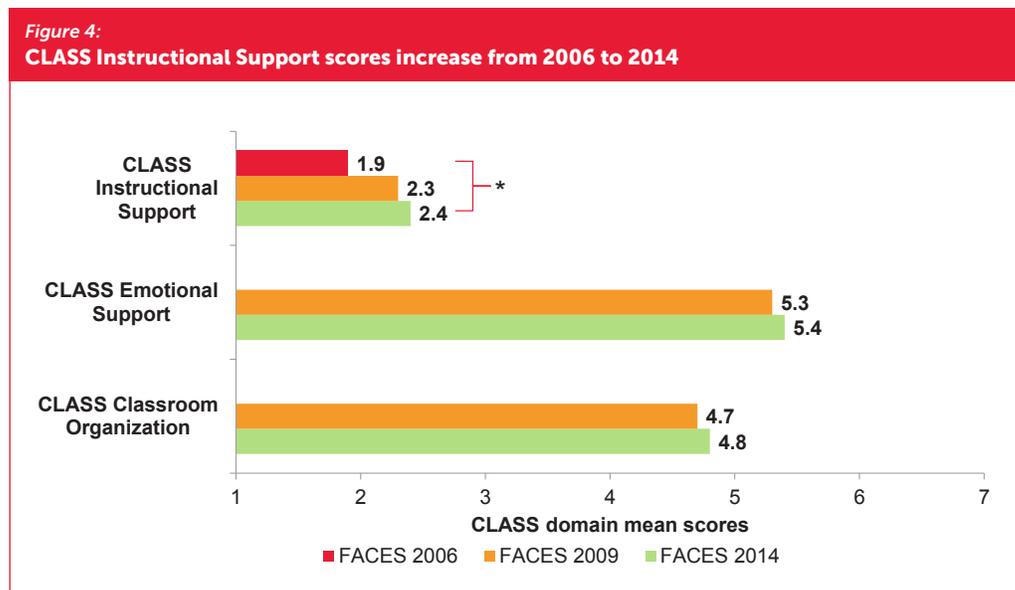
Source: Spring 2007, 2010, and 2015 FACES Classroom Observation.

Notes: Statistics are weighted to represent all Head Start classrooms in the cohort year.

Scores on the ECERS-R factors range from 1 to 7 with publisher-developed categories including inadequate (scores equal to 1 or 2), minimal (scores equal to 3 or 4), good (scores equal to 5 or 6), and excellent (scores equal to 7) quality. Data are drawn from Table B.1a in the accompanying technical report (Aikens et al. 2016).

\* Asterisk indicates that the change over time is statistically significant at the  $p \leq .05$  level.

Average Instructional Support scores on the Classroom Assessment Scoring System (CLASS; Pianta et al. 2008) improve from FACES 2006 to 2014, but do not change significantly from FACES 2009 to 2014. Similarly, there are no statistically significant changes in CLASS Emotional Support and Classroom Organization scores from FACES 2009 to 2014 (Figure 4).

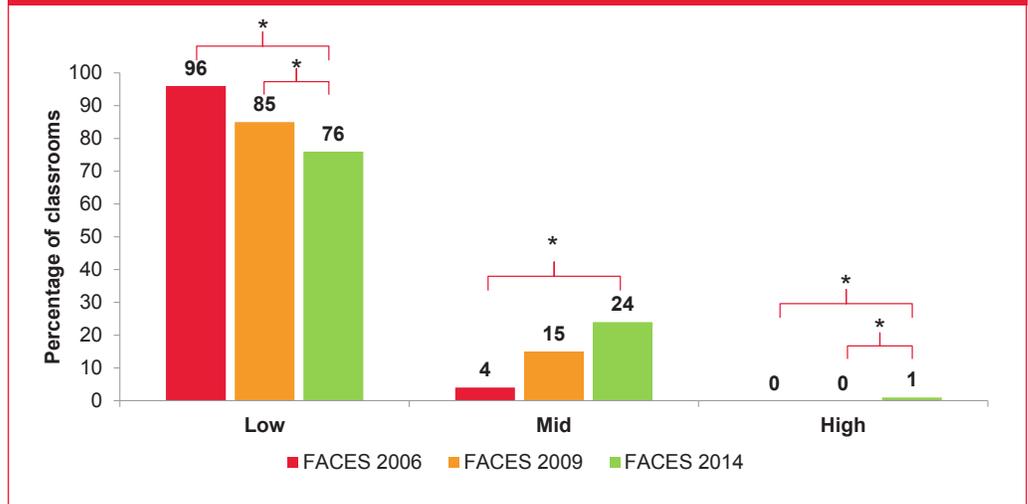


Source: Spring 2007, 2010, and 2015 FACES Classroom Observation.  
 Notes: Statistics are weighted to represent all Head Start classrooms in the cohort year. Only the Instructional Support domain of the CLASS is available in FACES 2006. Data are drawn from Table B.1 in the accompanying technical report (Aikens et al. 2016).  
 \* Asterisk indicates that the change over time is statistically significant at the  $p \leq .05$  level.

Similarly, based on publisher-developed categories, there are improvements in CLASS Instructional Support scores from FACES 2006 to 2014, with fewer classrooms scoring in the low range and more scoring in the mid or high ranges across cohorts (Figure 5).

From FACES 2009 to 2014 on the Instructional Support domain, fewer classrooms score in the low range, and more classrooms in the high range. On the Emotional Support domain, from 2009 to 2014, fewer classrooms score in the mid range, and more classrooms score in the high range (Figure 6). On the Classroom Organization domain, fewer classrooms score in the mid range between 2009 and 2014 but no statistically significant pattern of improvement is observed in the other categories of quality (Figure 7).

**Figure 5:**  
**Class Instructional Support: Classrooms in low range decrease from 2006 to 2014 and 2009 to 2014**



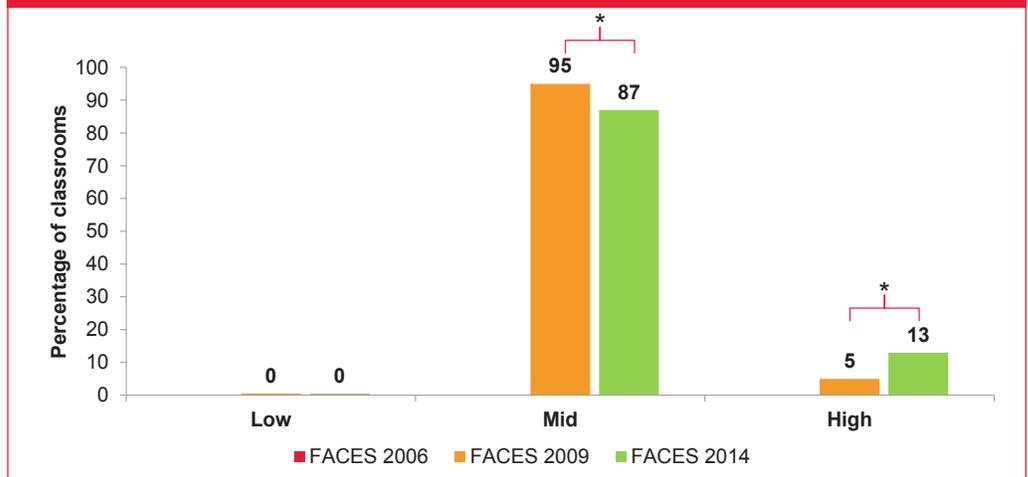
Source: Spring 2007, 2010, and 2015 FACES Classroom Observation.

Notes: Statistics are weighted to represent all Head Start classrooms in the cohort year.

Scores on the ECERS-R factors range from 1 to 7 with publisher-developed categories including inadequate (scores equal to 1 or 2), minimal (scores equal to 3 or 4), good (scores equal to 5 or 6), and excellent (scores equal to 7) quality. Data are drawn from Table B.1a in the accompanying technical report (Aikens et al. 2016).

\* Asterisk indicates that the change over time is statistically significant at the  $p \leq .05$  level.

**Figure 6:**  
**Class Emotional Support: Classrooms in high range increase from 2009 to 2014**



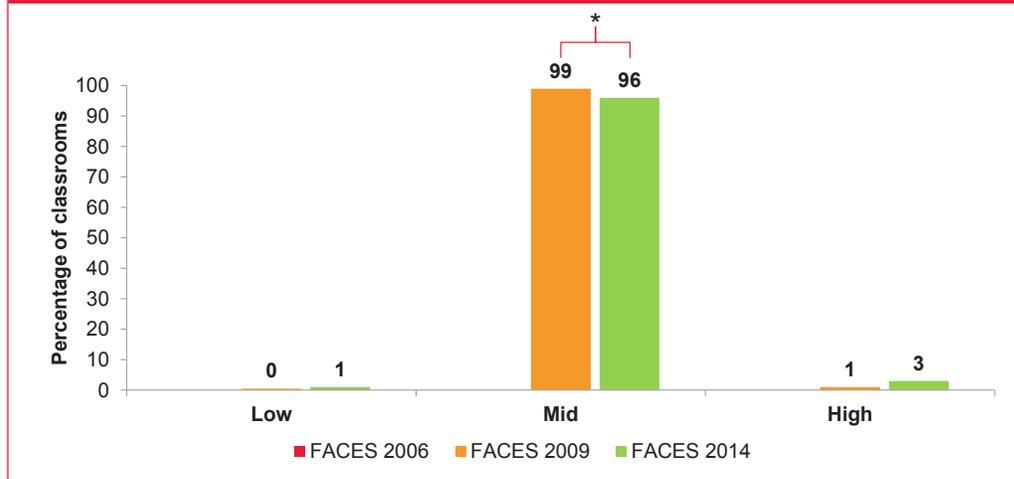
Source: Spring 2007, 2010, and 2015 FACES Classroom Observation.

Notes: Statistics are weighted to represent all Head Start classrooms in the cohort year.

Scores on the ECERS-R factors range from 1 to 7 with publisher-developed categories including inadequate (scores equal to 1 or 2), minimal (scores equal to 3 or 4), good (scores equal to 5 or 6), and excellent (scores equal to 7) quality. Data are drawn from Table B.1a in the accompanying technical report (Aikens et al. 2016).

\* Asterisk indicates that the change over time is statistically significant at the  $p \leq .05$  level.

**Figure 7:**  
**CLASS Classroom Organization: Most classrooms are in the mid range from 2009 to 2014**



Source: Spring 2007, 2010, and 2015 FACES Classroom Observation.

Notes: Statistics are weighted to represent all Head Start classrooms in the cohort year.

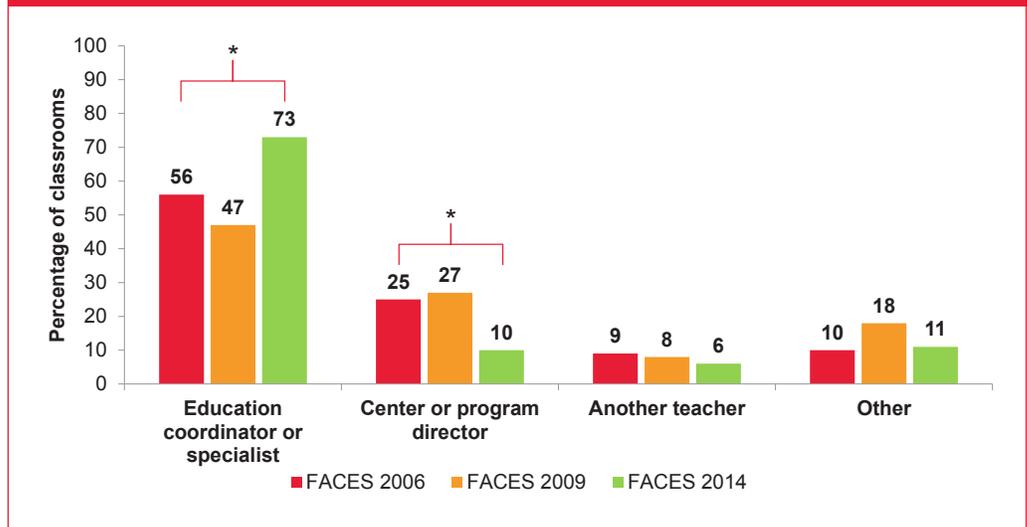
Scores on the CLASS domains range from 1 to 7 with publisher-developed categories including inadequate (scores equal to 1 or 2), minimal (scores equal to 3 or 4), good (scores equal to 5 or 6), and excellent (scores equal to 7) quality. Data are drawn from Table B.1a in the accompanying technical report (Aikens et al. 2016).

\* Asterisk indicates that the change over time is statistically significant at the  $p \leq .05$  level.

**Are there changes in mentoring in Head Start programs from FACES 2006 to 2014? Specifically, how many teachers report that they have a mentor? Who provides mentoring?**

We found that the prevalence of teacher mentoring remains stable from FACES 2006 to 2014, with about three-quarters of classrooms having teachers with a mentor, across cohorts. Among those classrooms in which teachers have mentors, there is a significant increase in teacher reports of mentoring by an educational coordinator or specialist and a significant decrease in mentoring by a center or program director from FACES 2006 to 2014 (Figure 8).

**Figure 8:**  
**Teacher mentoring by Education Coordinators increases from 2006 to 2014**



Source: Spring 2007, 2010, and 2015 FACES Teacher Interview/Survey.

Notes: Statistics are weighted to represent all Head Start classrooms in the cohort year.

The statistics presented are limited to classrooms in which teachers reported having mentors.

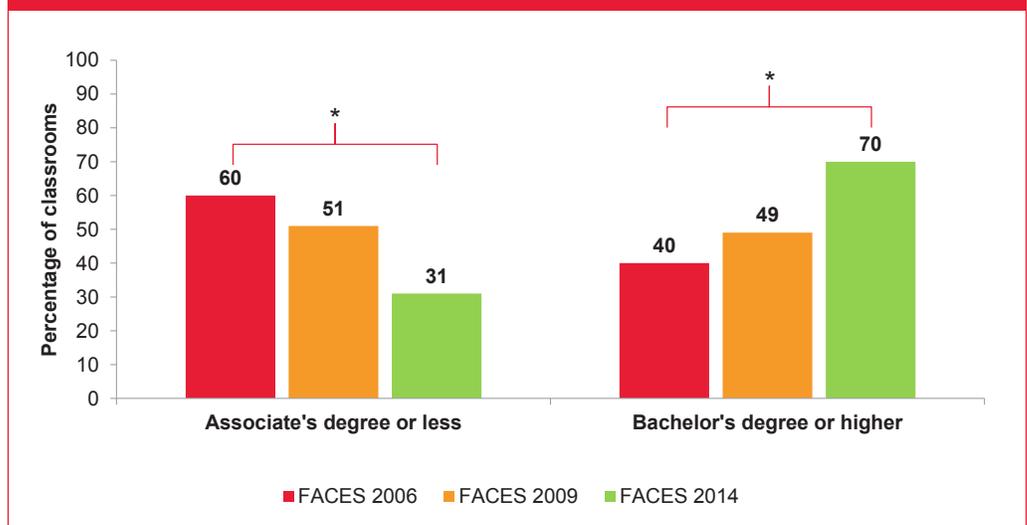
Data are drawn from Table B.2 in the accompanying technical report (Aikens et al. 2016).

\* Asterisk indicates that the change over time is statistically significant at the  $p \leq .05$ .

### Have the levels of education of Head Start teachers changed from FACES 2006 to 2014?

The percentage of classrooms that have a teacher with a bachelor’s degree or higher increases from FACES 2006 to FACES 2014 (Figure 9).

**Figure 9:**  
**Teachers increasingly have bachelor’s degrees from 2006 to 2014**



Source: Spring 2007, 2010, and 2015 FACES Teacher Interview/Survey.

Notes: Statistics are weighted to represent all Head Start classrooms in the cohort year.

Data are drawn from Table B.4 in the accompanying technical report (Aikens et al. 2016).

\* Asterisk indicates that the change over time is statistically significant at the  $p \leq .05$  level.

---

## **Are changes in observed classroom quality across cohorts explained by selected teacher characteristics (prevalence of mentoring, who provides mentoring, and level of education)?**

Finally, we examine whether changes in selected teacher characteristics (whether the teacher has a mentor and, if so, who serves as that mentor; and whether the teacher has at least a bachelor's degree) explain trends in observed classroom quality.<sup>3</sup> We hypothesize that changes in the selected teacher characteristics may account for some of the observed changes in classroom quality.

We analyze continuous ECERS-R factor scores and CLASS domain scores, as well as the distribution across publisher categories),<sup>4</sup> on the two ECERS-R factors and on CLASS Instructional Support and Emotional Support.<sup>5</sup> Analyses control for child demographic characteristics (average child age, percentage of children who are DLLs, and percentage of children with family income below poverty)<sup>6</sup> to account for the influence of changes in the demographic make-up of Head Start classrooms on observed classroom quality across cohorts.

Controlling for child demographic characteristics, the teacher characteristics included in the analyses:<sup>7</sup>

- Do not explain changes in average ECERS-R Teaching and Interactions or Provisions for Learning scores or the distribution across publisher-developed categories on these scores, from FACES 2006 to 2014 or from FACES 2009 to 2014.
- Partially explain changes in average CLASS Instructional Support scores from FACES 2006 to 2014. In fact, after controlling for child characteristics, they explain 15 percent of the change in these scores during this period. Specifically, whether the teacher has at least a bachelor's degree explains approximately 12 percent of the total classroom quality trend (that is, the increase in CLASS Instructional Support scores), while other characteristics in the model contribute more modestly. In short, the increase in teachers with at least a bachelor's degree accounts for a small portion of the improvement in CLASS Instructional Support from FACES 2006 to 2014.
- Do not explain trends in average Emotional Support or Classroom Organization scores, or in the distribution across publisher-developed categories for Instructional Support or Emotional Support.

---

## SUMMARY AND IMPLICATIONS

---

The goal of this brief was to examine trends in observed classroom quality and selected teacher characteristics across FACES cohorts. We also examined whether changes in selected teacher characteristics are related to changes in observed classroom quality.

We see increases in average classroom quality from FACES 2006 to 2014 on both ECERS-R factors and on CLASS Instructional Support. From FACES 2009 to 2014, there are increases on average ECERS-R factor scores, but not on CLASS domain scores. In examining publisher categories, we found more classrooms score in the good or excellent range on both ECERS-R factors from 2006 to 2014, but from 2009 to 2014 only on ECERS-R Provisions for Learning scores. Fewer classrooms score in the low range on CLASS Instructional Support from 2006 to 2014. From 2009 to 2014, fewer score in the low range on Instructional Support, and more score in the high range on Emotional Support.

Across this period of time we found changes in some teacher characteristics, also trending in a positive direction. However, changes in the teacher characteristics that we examined account for only 15 percent of the trends in observed classroom quality. Specifically, of the variables we examined, whether the teacher has at least a bachelor's degree explains approximately 12 percent of the trend in only average CLASS Instructional Support scores. The change in mentors from directors to education coordinators does not appear to account for changes in observed quality. Other factors might help to explain changes in observed quality, including those that are not measured in FACES. For example, although FACES captures information on the prevalence of mentoring and the staff who provide mentoring to teachers, the quality, intensity, and intentionality of that mentoring may be more critical for explaining improvements in classroom quality over time.

Future work might examine additional factors that could be associated with observed classroom quality, such as program management quality or the use of data for quality improvement. Identifying processes that can explain the recent increases in Head Start classroom quality may help target resources for quality improvement efforts in the future. Additionally, further research would be needed to establish causal relationships between quality improvement efforts, changes in teacher characteristics, and changes in observed classroom quality.

### Box 1: The FACES design across cohorts

The FACES sample provides information at the national level about Head Start programs, centers, and classrooms, and the children and families they serve. FACES 2006 and 2009 focused on newly entering children and their classrooms and programs; FACES 2014 represented all children—whether newly entering or returning—and their classrooms and programs.

Information on teacher characteristics reported in this brief draw from interviews or surveys teachers completed about their classrooms and themselves. We also use information from parent interviews or surveys to describe children’s demographic characteristics. To measure the quality of Head Start classrooms, analyses use the Classroom Assessment Scoring System (CLASS) and the Early Childhood Environment Rating Scale–Revised (ECERS-R). Both the CLASS and the ECERS-R items are scored on a seven-point scale, with higher scores reflecting better quality care. Table 1 describes the factors and domains measured by the ECERS-R and the CLASS. The content of the quality observations varies across FACES cohorts. For example, only the Instructional Support domain of the CLASS is available in FACES 2006. However, all three domains of the CLASS are available in later cohorts. Similarly, whereas FACES 2006 used the full ECERS-R, FACES 2009 and 2014 used only a short form with an abbreviated set of items based on findings in other large-scale studies (Clifford et al. 2005). More details on the study methodology and response rates in each cohort appear in the user’s manual that accompanies the public use data files (West et al. 2010; Malone et al. 2013; Kopack Klein et al. 2016).

**Table 1:**  
**Description of ECERS-R factors and CLASS domains**

Factors or domains	Characteristics assessed by factors or domains
ECERS-R Teaching and Interactions	Quality of teacher-child interactions
ECERS-R Provisions for Learning	Materials available in the classroom and the arrangement of classroom space
CLASS Instructional Support	Quality of instructional practices used in the classroom
CLASS Emotional Support	Social and emotional functioning in the classroom
CLASS Classroom Support	Teacher’s ability to organize the classroom to make efficient use of class time

## Box 2: Analytic approach

The analyses proceeded in stages:

- First, we conducted descriptive analyses of measures of classroom quality, teacher characteristics, and child demographic characteristics (for example, means and percentages). The full set of analyses examined observed classroom quality, group size, and child-adult ratios; child demographic characteristics; curricular supports; teacher participation in training and workshops; prevalence and frequency of mentoring and by whom; and teacher education and credentials, depressive symptoms, job satisfaction, and turnover. The current brief presents findings from only a subset of these analyses that met certain criteria, as detailed below. Details on patterns for all characteristics are described in the accompanying technical report (Aikens et al. 2016).
- Second, we conducted regression analyses to examine trends over time (2006 to 2014 and 2009 to 2014) for each measure of classroom quality and the characteristics listed above.
- Third, based on the trend analyses, we then conducted mediation analyses to examine the characteristics that might explain or affect trends in the key outcome of interest—observed classroom quality. Specifically, we included the following teacher characteristics as mediators in the models: education (bachelor’s degree or higher versus less than a bachelor’s degree), and mentoring (whether there was a mentor and who provided the mentoring versus no mentoring). We also included child demographic characteristics—average child age, percentage of children in the classroom who were dual language learners, and percentage of children in the classroom with family incomes below poverty—as control variables in the models. We selected the control variables a priori and the mediators based on their performance in the trend analyses (selecting variables that changed significantly across cohorts or changed in ways likely to explain change in observed quality) and for theoretical reasons (selecting variables that we hypothesized would explain improvements in classroom quality or would be influenced by Head Start policies). In the accompanying technical report (Aikens et al. 2016), we provide more information on how we coded and included each of the control variables and mediators in the models.

In both the trend and mediation analyses, we used linear regressions for continuous variables and simple logistic regressions for dichotomous variables. Each model included a single variable of interest as the predicted variable (for example, CLASS Instructional Support mean scores) and dummy codes for cohort (for example, 2009, 2014) as the predictor variables to measure the trend or change in quality across cohorts. In the mediation models, we added the control and mediating variables to examine whether the trends in quality persisted after accounting for them.

The accompanying technical report includes more details on the analyses conducted (Aikens et al. 2016). All analyses were weighted, with the appropriate strata and primary sampling unit variables applied, to account for the multistage sample design (Kopack Klein et al. 2016).

---

## REFERENCES

- Aikens, Nikki, Charles Bush, Philip Gleason, Lizabeth Malone, and Louisa Tarullo. "Tracking Quality in Head Start Classrooms: FACES 2006 to FACES 2014 Technical Report." Report submitted to the U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation. Washington, DC: Mathematica Policy Research, October 2016.
- Clifford, Richard, Oscar Barbarin, Florence Chang, Diane M. Early, Donna Bryant, Carollee Howes, Margaret Burchinal, and Robert Pianta. "What Is Pre-Kindergarten? Characteristics of Public Pre-Kindergarten Programs." *Applied Developmental Science*, vol. 9, no. 3, 2005, pp. 126–143.
- Harms, Thelma, Richard Clifford, and Debby Cryer. "Early Childhood Environment Rating Scale–Revised (ECERS-R)." New York, NY: Teachers College Press, 1998.
- Kopack Klein, Ashley, Barbara Lepidus Carlson, Nikki Aikens, Anne Bloomenthal, Jerry West, Lizabeth Malone, Emily Moiduddin, Melissa Hepburn, Sara Skidmore, Sara Bernstein, Annalee Kelly, Felicia Hurwitz, and Grace Lim. "Head Start Family and Child Experiences Survey (FACES 2014) User's Manual." Draft report submitted to the U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation. Washington, DC: Mathematica Policy Research, October 2016.
- Malone, Lizabeth, Barbara Lepidus Carlson, Nikki Aikens, Emily Moiduddin, Ashley Kopack Klein, Jerry West, Annalee Kelly, Cassandra Meagher, Anne Bloomenthal, Lara Hulse, and Kristina Rall. "Head Start Family and Child Experiences Survey: 2009 User's Manual." Report submitted to the U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation. Washington, DC: Mathematica Policy Research, April 2013.
- Pianta, R., K. LaParo, and B. Hamre. "The Classroom Assessment Scoring System Pre-K Manual." Charlottesville, VA: University of Virginia, 2008.
- West, Jerry, Nikki Aikens, Barbara Carlson, Cassandra Meagher, Lizabeth Malone, Anne Bloomenthal, Annalee Kelly, Kristina Rall, and Rita Zota. "Head Start Family and Child Experiences Survey: 2006 User's Manual." Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation, August 2010.

---

## ENDNOTES

- <sup>1</sup> Throughout this brief, we use the term teacher to refer to the lead teacher working in the classroom with children.
- <sup>2</sup> Given the analyses focus on changes in observed quality as measured by the Early Childhood Environment Rating Scale–Revised (ECERS-R; Harms et al. 1998) and the Classroom Assessment Scoring System (CLASS; Pianta et al. 2008), we focus on the period between 2006 and 2014, during which these measures were available in FACES.
- <sup>3</sup> As noted in Box 2, the full set of analyses examined change in other possible mediators available in FACES. This includes curricular supports; teacher participation in training and workshops; and teacher education and credentials, depressive symptoms, job satisfaction, and turnover. We excluded most of these characteristics as mediators from the analyses because they did not change over time. A handful of other characteristics changed over time but were not included as mediators: ongoing supports for curriculum use, the staff providing teachers with ongoing supports for curriculum use, and mean level of teacher depressive symptoms. With regard to supports related to curriculum use, we found that there was a decrease in some types of support over time, and we would not expect that such a decrease would be associated with an increase in quality. On teacher depressive symptoms, while we saw a small significant decline in level of depressive symptoms over time, that shift was not accompanied by a change in the percentage of teachers who meet criteria for depression. For all of these reasons, we excluded these variables from the mediation models. Details on the patterns for all characteristics are described in the accompanying technical report (Aikens et al. 2016).
- <sup>4</sup> The analyses focused on continuous quality scores provide information on whether the selected teacher characteristics explain changes in average scores over time, whereas those that focus on the distribution across publisher categories identify whether the selected teacher characteristics explain changes in the percentage of classrooms meeting different quality categories over time (for example, with "low" Instructional Support scores).

<sup>5</sup> We omitted CLASS Classroom Organization categories from these analyses because the scores on this domain were unevenly distributed. This resulted in nearly all classrooms (over 96 percent in both 2009 and 2014) falling into the same category, and fewer than 10 classrooms falling in the high category or in the low category in each year. So while the change in the percentage of classrooms falling in the middle category was statistically significant, we felt that this change was not of practical importance and thus not useful to include in this analysis.

<sup>6</sup> We aggregated child demographic characteristics up to the classroom level from parent interview and survey data. As a result, the characteristics reflect the characteristics of sampled children rather than those of all children in the classrooms. The child sample included 10 to 12 children per classroom, out of an average observed class size of 14 to 15 children.

<sup>7</sup> We derive an estimate of the proportion of the trend in classroom quality that is explained by the control variables (child characteristics) alone and the controls and mediators (teacher characteristics) combined. We highlight only those instances in which the controls or mediators explain a substantial proportion of the trend—in these instances, we used a threshold of 10 percent (that is, the time or cohort regression coefficient changes by more than 10 percent). A change in the regression coefficient of less than 10 percent would suggest that most of the trend in observed quality (90 percent) is driven by factors other than the controls and mediators included in the models.