Massachusetts Preschool Expansion Grant (PEG) Evaluation: Spring 2018 Statewide PEG Classroom Observation Report

In the spring of 2018 (March-April), data collectors trained by Abt Associates staff conducted an observation of each PEG classroom using the Classroom Assessment Scoring System (CLASS) and the Early Language and Literacy Classroom Observation (ELLCO). The observation began when the school day started and continued until children started naptime (typically 8 a.m. – 1 p.m.).

The **CLASS** measures **overall instructional quality** with a focus on interactions among teachers and students in the classroom. The CLASS produces four scores, an overall score and a score for each of three domains. Each score ranges from 1 to 7. A score of 1-2 is described as "low" quality in that aspect of teacher-child interaction. Scores of 3-5 are described as "moderate," and scores of 6-7 are described as "high" quality.

- ***Total Score***: The total score combines the three domains.
- ***Instructional Support Domain***: How well the teachers use instruction to help children learn. The domain has 3 dimensions: Concept Development, Quality of Feedback, and Language Modeling.
  - This year, in order to provide more detailed formative feedback about the types of instructional supports that were observed, we collected additional more nuanced information that is typically not recorded with the CLASS. Observers used a 3-point scale to rate each individual indicator in the 3 Instructional Support dimensions for each 20-minute cycle. We also collected information related to the context of each 20-minute cycle to provide feedback related to quality in various groupings and with various content focus.
- ***Emotional Support Domain:*** How well teachers promote a positive climate through their interactions. The domain has 4 dimensions: Positive Climate, Negative Climate, Teacher Sensitivity, and Regard for Student Perspectives.
- ***Classroom Organization Domain***: How well the classroom is organized to facilitate learning. The domain includes 3 dimensions: Behavior Management, Productivity, and Instructional Learning Formats.

The **ELLCO** measures the **quality of support for language and literacy development**, including support for diversity of languages, abilities, and cultures. The ELLCO produces three scores, an overall score and a score for each of two subscales. Each score ranges from 1 to 5. A score of 1 is described as "deficient" evidence of quality. A score of 2 is described as "inadequate," a score of 3 is described as "basic," a score of 4 is described as "strong," and a score of 5 is described as "exemplary."

- ***Total Score***: The total score combines the two subscales.
- ***General Classroom Environment Subscale***: How well the classroom is organized and promotes general learning. The subscale has 2 domains: Classroom Structure and Curriculum.
- ***Language and Literacy Subscale***: How teachers and students engage in literacy and language activities. The subscale has 3 domains: Language Environment, Books and Book Reading, and Print and Early Writing.

How to Interpret the Data

On the pages that follow, you will see 7 pages of data for each measure. On the first page of data of each measure's summary (Figures 1 and 6), you will see several column graphs. Each column represents the *average score* across classrooms in each community. Each of the five communities is represented by a different colored column: Boston's data is in dark blue, Holyoke's is in orange, Lawrence's is in green, Lowell's is in red, and Springfield's is in purple. The light blue columns show average scores for the *PEG classrooms across Massachusetts.* These graphs should show you how similarly or differently the communities scored on average in comparison to the 48 PEG classrooms across Massachusetts.

On the second page of each measure's summary (Figures 2 and 7), you will see arrays of circles. Each panel or row contains either total scores or subscale scores. These graphs show you the individual scores of each Massachusetts PEG classroom observed. The color scheme for these graphs is the same as the one used and described above. Each panel also has a dotted gray line, representing the average score on that subscale across the state.

On the third page of each measure's summary (Figure 3), you will see several column graphs showing change over time on each measure for the state overall and for each community, along with an indication of which Year 1-Year 3 changes are statistically significant in the full state sample.

On the fourth page of each measure's summary (Figures 4-5), you will see bar graphs that represent the percentage of classrooms that fell into different quality categories, using categories as constructed by the measure developers, in each year. These graphs should give you an indication of the levels of classroom quality from Winter 2018, Winter 2017, and Winter 2016 to indicate how quality has changed over time. Because of teacher turnover in 31% of the classrooms (15 of the 48), the first bar graph represents all 48 classrooms, and the second bar graph is limited to those 33 classrooms in which the same lead teacher was present for observations in all three years.

The last section of this report describes the CLASS Instructional Support domain, specifically, in greater detail and provides data that can be used for formative feedback for programs.

Important to Remember: There are some important things to keep in mind when considering these data:

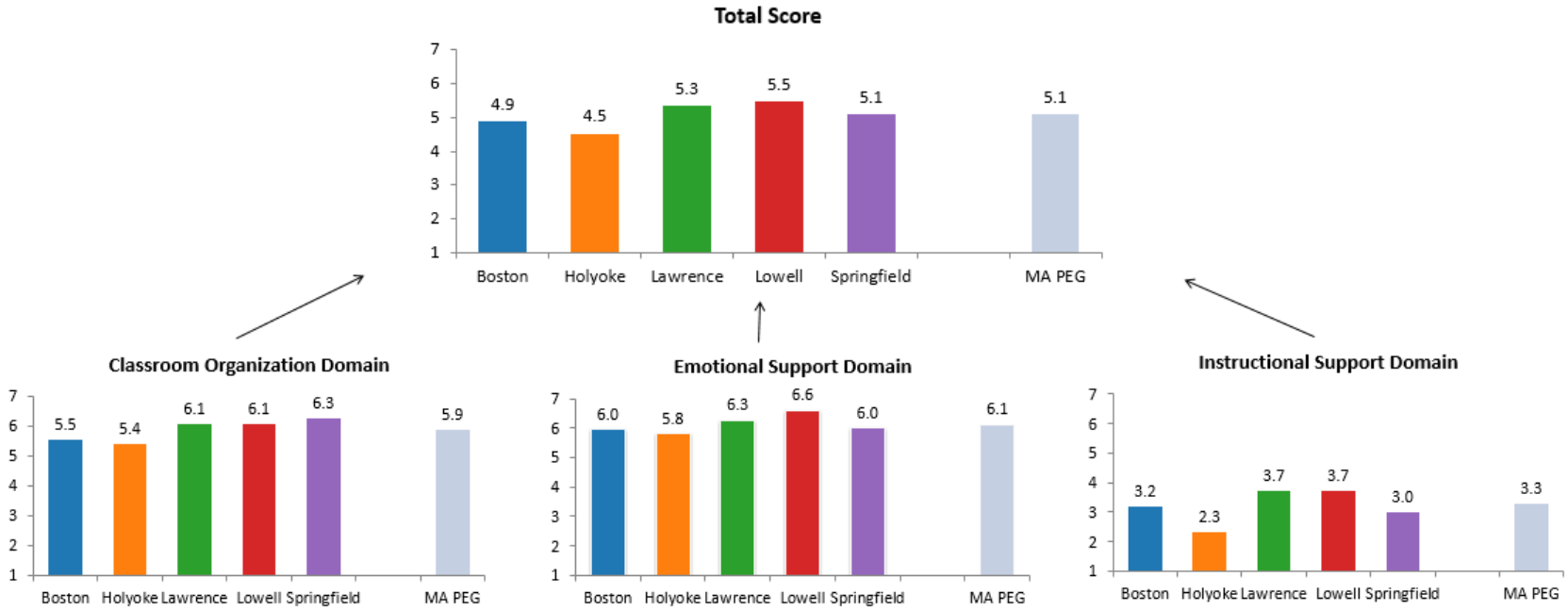
- These data come from a single day of observation only. It is likely that a classroom changes slightly from one day to the next. However, the information presented should be a relatively representative view of the classrooms.
- These data should not be used to evaluate classroom effectiveness, but rather to provide a starting point to examine where programs are exceling and where there might be room for improvement.

CLASS Scores for PEG Classrooms in Year 3 of PEG Implementation (2017-18)

Figure 1 shows the average CLASS scores for PEG statewide and by community in the third year of PEG. The average scores for Classroom Organization and Emotional Support are at or near what is considered to be a high level of quality, while the average score for Instructional Support is at the low end of the range of scores considered to be moderate quality.

Figure 2 shows the CLASS scores for each PEG classroom in relationship to the average state score, organized by community.

Figure 1: Average Scores for CLASS Total and Domain Scores Statewide and by PEG Community, Spring 2018



Notes. MA PEG statewide sample=48 classrooms (Boston=15, Holyoke=4, Lawrence=10, Lowell=8, Springfield=11). Each graph shows the average score across classrooms on the CLASS 1-7 scale. For example, the top graph shows that, on average, PEG classrooms across Massachusetts scored 5.1 out of 7 points on the total CLASS score, with 7 representing the highest observed quality.

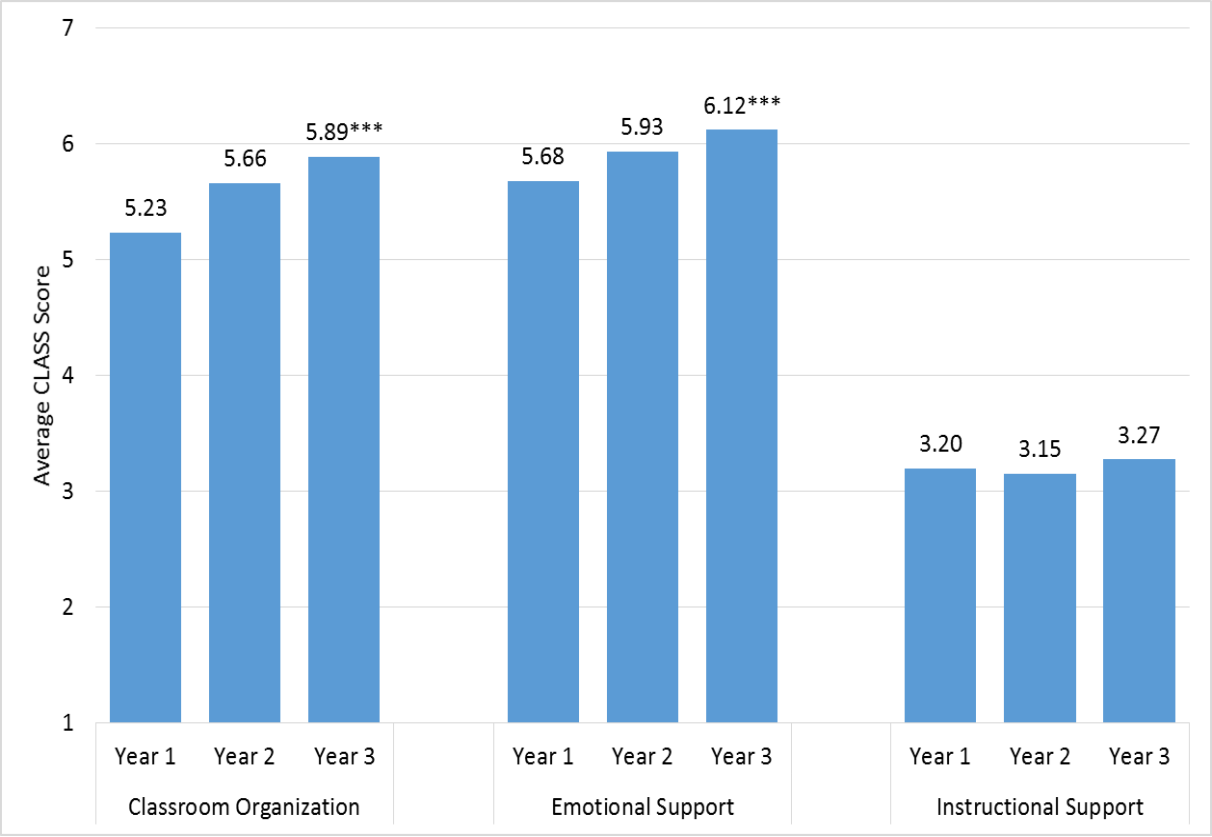
Figure 2: Average Scores for CLASS Total and Domain Scores by Community and Classroom, Spring 2018

Notes. Each circle represents an individual classroom's score on the CLASS 1-7 scale, and the dotted line shows the average score for the state.

Change Over Time in CLASS Scores

Figure 3 below shows the mean CLASS domain scores over time for the full sample of 48 classrooms. Significant change from Year 1 to Year 3 is indicated on the Year 3 bars with asterisks. PEG classrooms improved their CLASS scores significantly from the first to the third year of the program for both the Classroom Organization and Emotional Support domains. Though Instructional Support scores did improve from Year 1 to Year 3, the change was not statistically significant.

Figure 3: Change over Time in CLASS Scores for State, All 48 Classrooms

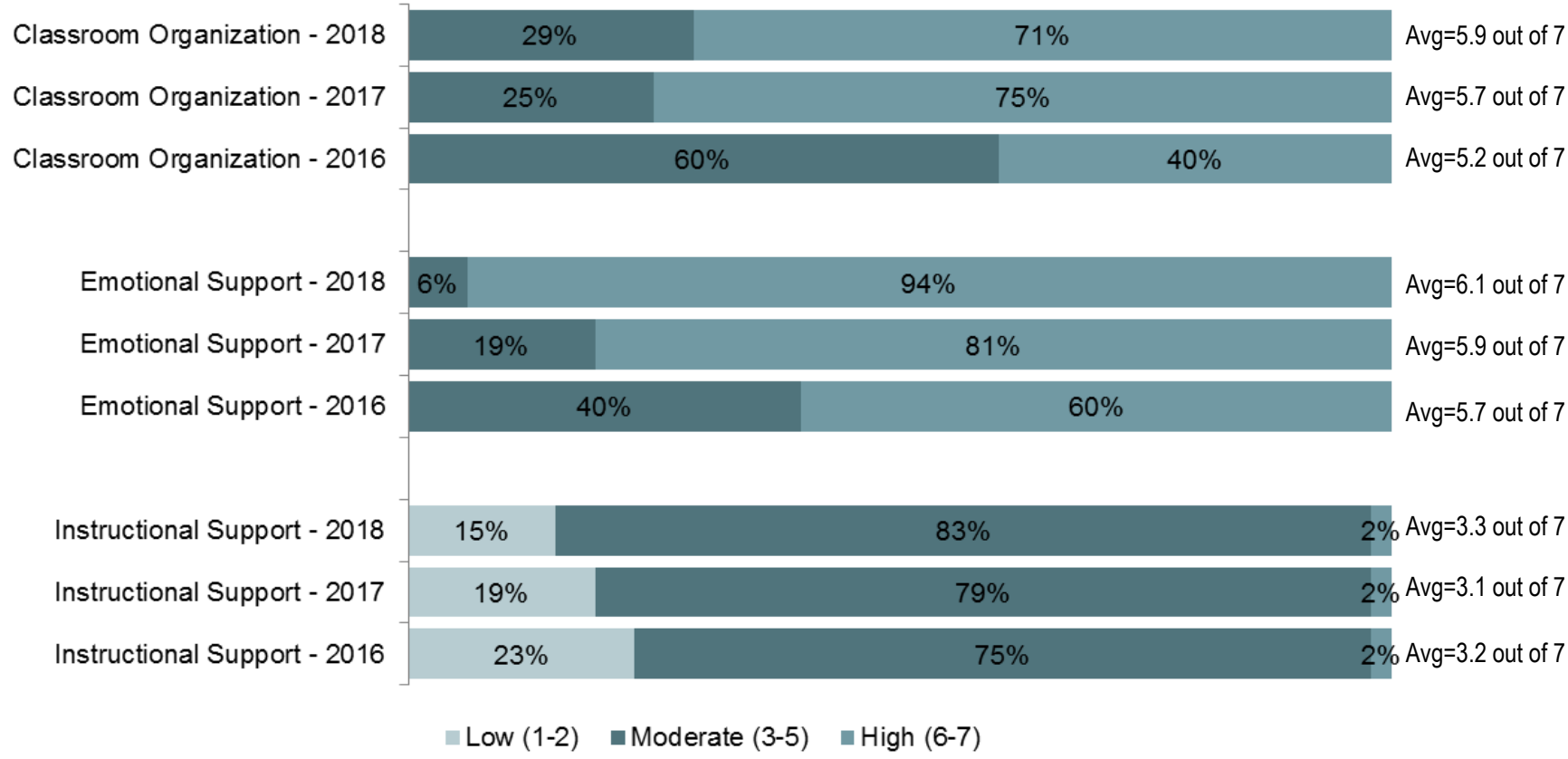


***p<.001

Another way to examine the change over time in CLASS scores is by creating levels of quality based on CLASS scores, with Low Quality defined as a score of 1 or 2, Moderate Quality as a score of 3-5, and High Quality as a score of 6 or 7. Using these levels, the change over time can be described in terms of the percentage of classrooms in each year scoring at each of these three levels of quality. Figure 4 shows the percentages of all classrooms at each level of quality in each year. The pattern of change over time differed for the three domains. For the Classroom Organization and Emotional Support domains, the percentage of classrooms scoring at a high level of quality changed markedly from year 1 to year 2. For Emotional Support, there was continued marked improvement between year 2 and year 3; by year 3, almost all classrooms were at a high level of quality. For Classroom Organization, the growth from year 1 was sustained but did not continue to increase from year 2 to year 3. For the Instructional Support domain, there was no change from year 1 to year 2 in the percentage of classrooms with high quality.

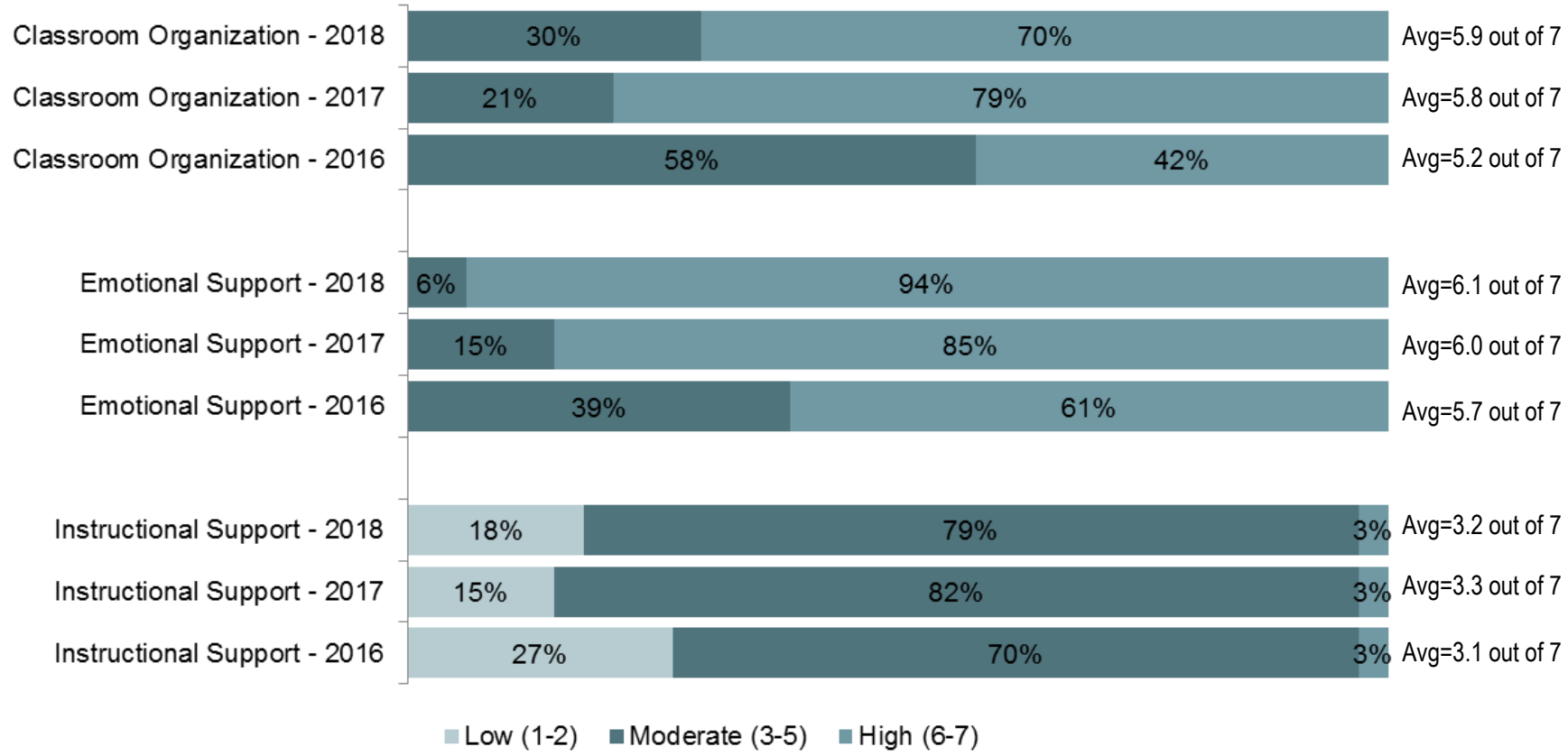
However, there was a decreased percentage of classrooms with low quality over the three years and an increase in the percentage of classrooms at a moderate level of quality. When the same analysis is conducted with the subset of 33 classrooms with the same lead teacher across the three years, a similar pattern is observed for all three CLASS domains (Figure 5).

Figure 4: Change over Time in Percentage of PEG Classrooms in Quality Categories by CLASS Domain, All 48 Classrooms



Notes. Each bar shows the percent of the 48 PEG classrooms (Boston=15, Holyoke=4, Lawrence=10, Lowell=8, Springfield=11) that fell into the different CLASS quality ranges according to their observed scores in 2016, 2017, and 2018. For example, the bottom three bars show that 85% of Massachusetts PEG classrooms scored in the 'Moderate' or 'High' quality category on Instructional Support in 2018, compared to 81% in 2017 and 77% in 2016.

Figure 5: Change over Time in Percentage of PEG Classrooms in Quality Categories by CLASS Domain, 33 Classrooms with Same Teacher All 3 Years



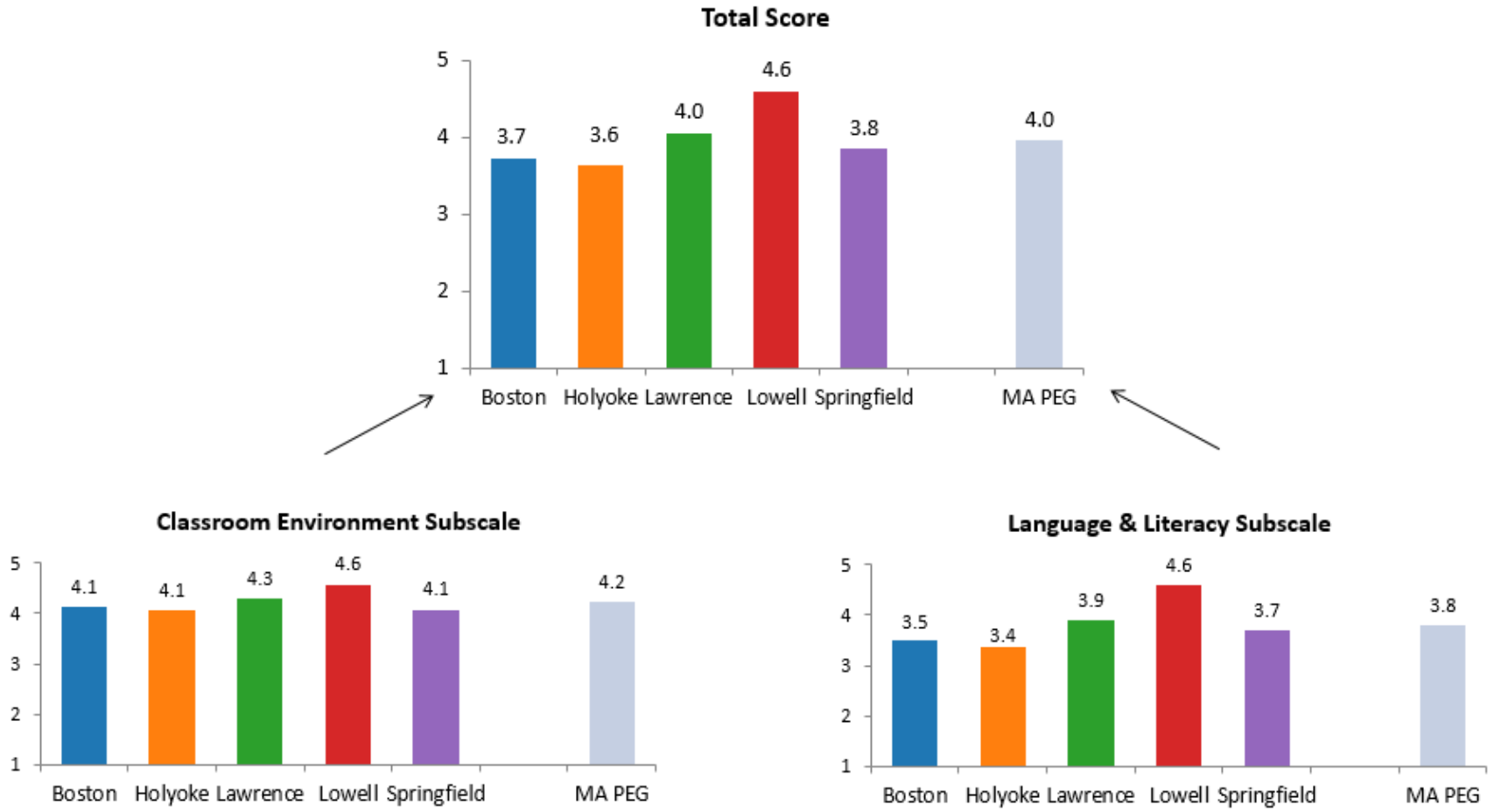
Notes. Each bar shows the percent of the 33 PEG classrooms (Boston=9, Holyoke=3, Lawrence=8, Lowell=5, Springfield=8) with the same lead teacher in the first three years of PEG that fell into different CLASS quality ranges according to their observed scores in 2016, 2017, and 2018. For example, the bottom three bars show that 82% of those 33 classrooms scored in the 'Moderate' or 'High' quality category on Instructional Support in 2018, compared to 85% in 2017 and 73% in 2016.

ELLCO Scores for PEG Classrooms in Year 3 of PEG Implementation (2017-18)

Figure 6 shows the average ELLCO scores for PEG statewide and by community in the third year of PEG. The average scores for both the Classroom Environment and Language and Literacy subscales are at what is considered to be a strong level of quality. Looking at the five PEG communities, the pattern across the two subscales by community is similar to the pattern statewide, although there are differences in the level of quality achieved. For Classroom Environment, the average scores for the five communities are all in the range of strong quality. For Language and Literacy, one community has an average score that is at the high end of strong quality, and the other four communities are at the range of scores considered to be basic quality.

Figure 7 shows the ELLCO scores for each PEG classroom in relationship to the average state score, organized by community.

Figure 6: Average ELLCO Total and Subscale Scores Statewide and by PEG Community, Spring 2018



Notes. Statewide sample=48 classrooms (Boston=15, Holyoke=4, Lawrence=10, Lowell=8, Springfield=11). Each graph shows the average score across classrooms on the ELLCO 1-5 scale. For example, the top graph shows that, on average, PEG classrooms across Massachusetts scored 4.0 out of 5 points on the total ELLCO score, with 5 representing the highest observed quality.

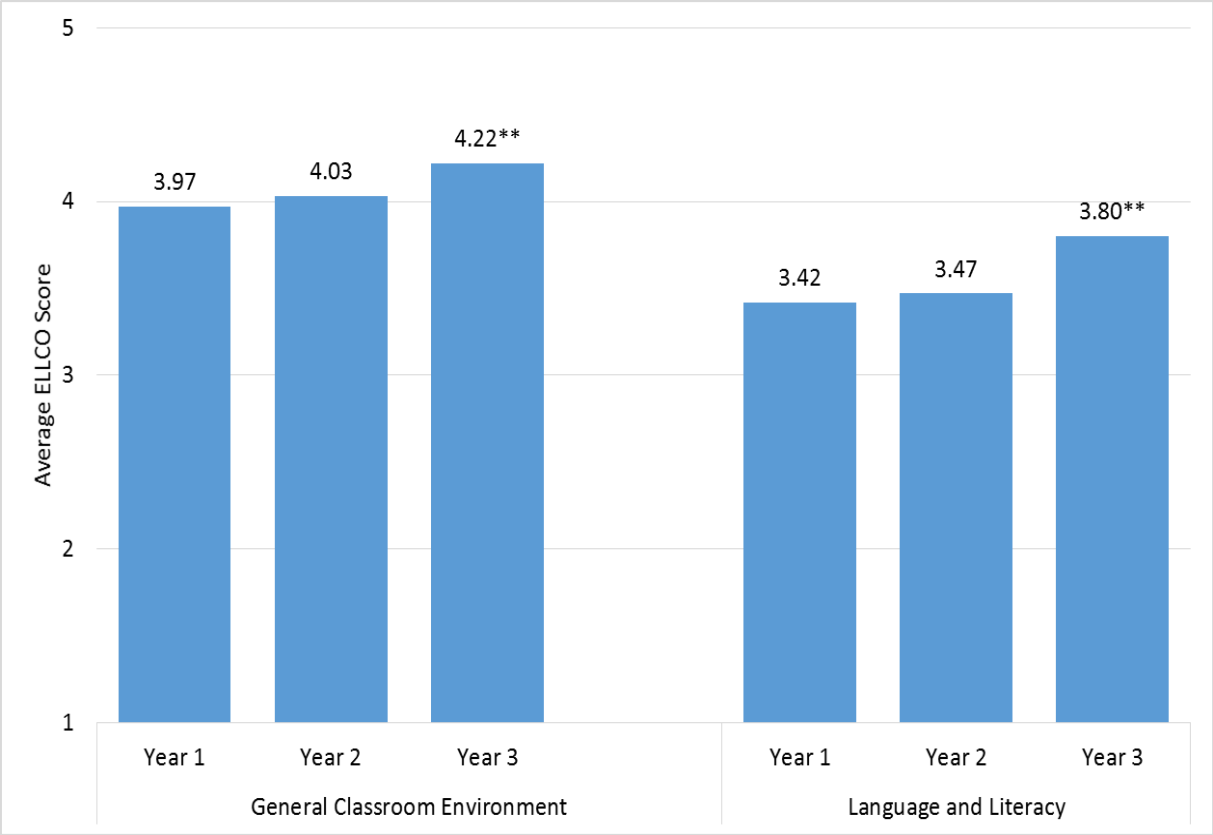
Figure 7: Average ELLCO Total and Subscale Scores by Community and Classroom, Spring 2018

Notes. Each circle represents an individual classroom's score on the ELLCO 1-5 scale, and the dotted line shows the average score for the state.

Change Over Time in ELLCO Scores

Figure 8 shows the mean ELLCO subscale scores over time for the full sample of 48 classrooms. Significant change from Year 1 to Year 3 is indicated on the Year 3 bars with asterisks. PEG classrooms improved their ELLCO scores significantly from thefirst to the third year of PEG on both subscales.

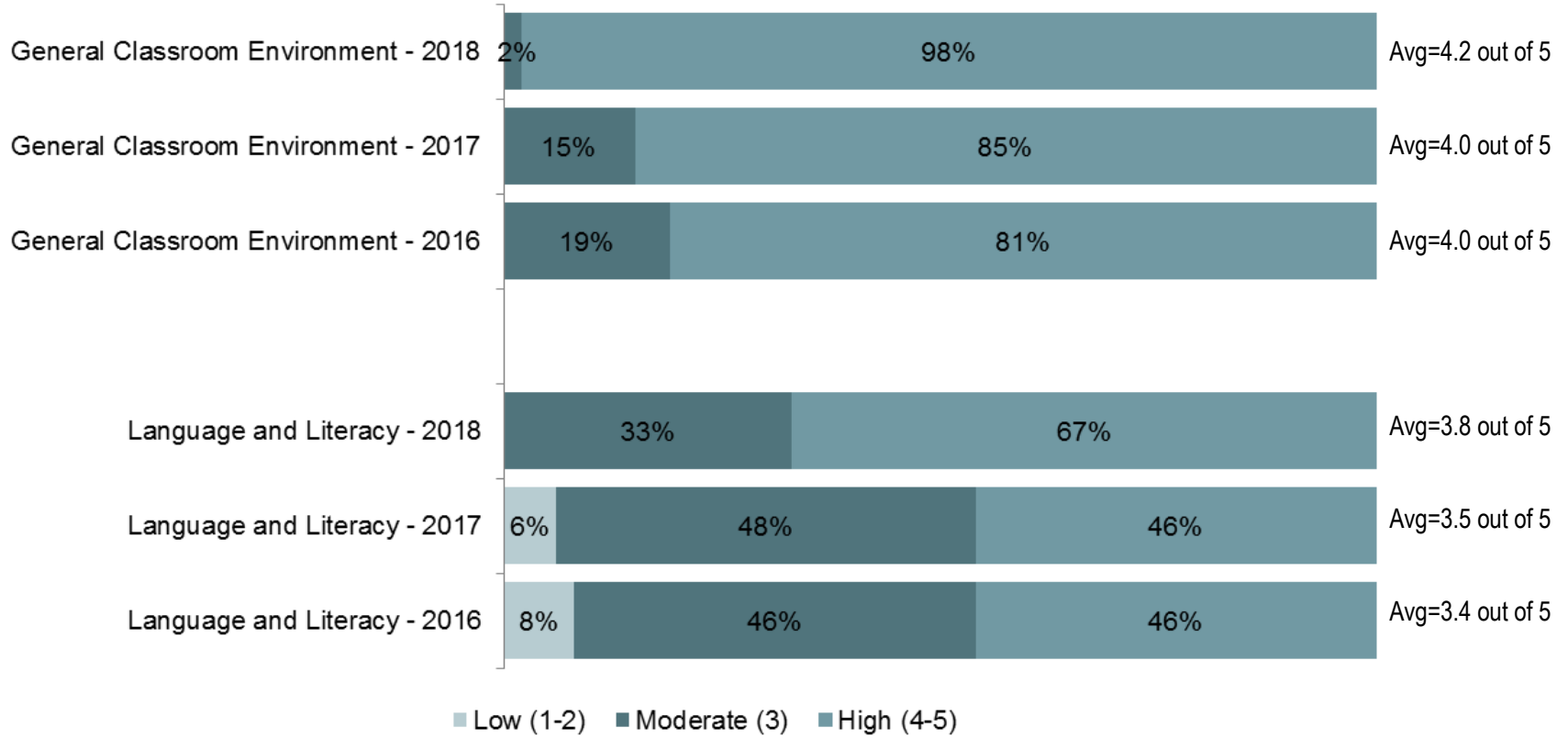
Figure 8: Change over Time in Average ELLCO Scores for State, All 48 Classrooms



**p<.01

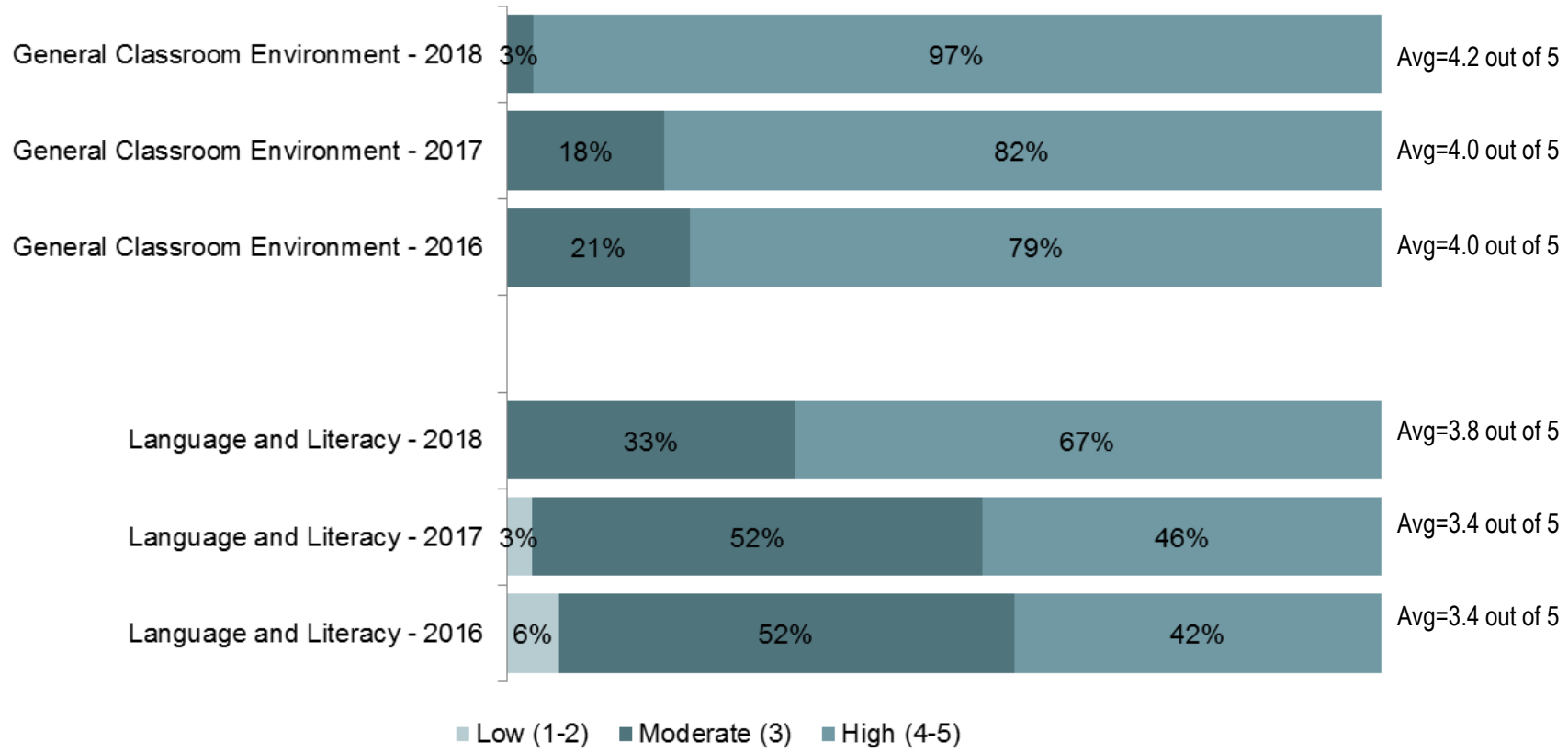
Change over time in ELLCO scores was also examined in terms of the level of quality based on the ELLCO scores, with Low Quality defined as a score of 1 or 2, Moderate Quality as a score of 3, and High Quality as a score of 4 or 5. Using these levels, the change over time can be described in terms of the percentage of classrooms in each year scoring at each of these three levels of quality. Figure 9 shows the percentages of all classrooms at each level of quality in each year. The pattern of change over time differed for the two subscores. For the General Classroom Environment subscale, by the end of year 3, nearly all of the classrooms were at a high level of quality. There was a small increase in scores from year 1 to year 2 and a larger improvement from year 2 to year 3. For the Language and Literacy subscale, most of the increase occurred between years 2 and 3; in that time, all of the programs that were at a low level of quality improved to at least a moderate level, and the percentage of classrooms at a high level of quality increased by nearly 20 percent. The pattern of change was similar in the full sample of 48 classrooms and the subset of 33 classrooms where the teacher was the same across all three years (Figure 10).

Figure 9: Change over Time in Percentage of PEG Classrooms in Quality Categories by ELLCO Subscale, All 48 Classrooms



Notes. Each bar shows the percent of the 48 PEG classrooms (Boston=15, Holyoke=4, Lawrence=10, Lowell=8, Springfield=11) that fell into ELLCO quality ranges according to their observed scores in 2016, 2017, and 2018. For example, the bottom three bars show that 100% of Massachusetts PEG classrooms scored in the 'Moderate' or 'High' quality category on Language and Literacy in 2018, compared to 94% in 2017 and 92% in 2016.

Figure 10: Change over Time in Percentage of PEG Classrooms in Quality Categories by ELLCO Subscale, 33 Classrooms with Same Teacher in 2016 and 2017



Notes. Each bar shows the percent of the 33 PEG classrooms (Boston=9, Holyoke=3, Lawrence=8, Lowell=5, Springfield=8) with the same lead teacher in the first three years of PEG that fell into ELLCO quality ranges according to their observed scores in 2016, 2017, and 2018. For example, the bottom three bars show that 100% of those 33 classrooms scored in the 'Moderate' or 'High' quality category on Language and Literacy in 2018, compared to 97% in 2017 and 94% in 2016.

How Individual Instructional Support Indicators Relate to the Overall CLASS Instructional Support Score

In PEG, as is true in most other samples of preschool classrooms, scores on Instructional Support are the lowest of the three CLASS domains. This is of particular concern because there is evidence that children's academic skills are more strongly related to scores on Instructional Support than they are to the scores on the other two CLASS domains (Emotional Support and Classroom Organization).

To try to generate more helpful information for teachers from the CLASS observations, we looked "underneath" the overall scores on Instructional Support to examine the subscores that observers assigned during the observation period that made up the total score. For the Instructional Support domain, observers considered fourteen specific instructional behaviors (termed "Indicators" in the CLASS manual) within three dimensions (see table below). Typically, observers are not required to formally record ratings at the indicator level. For this round of observations, coders were instructed to enter indicator ratings on a **3-point scale (high/medium/low)** to see if these data might provide useful information to ELPs and teachers about where practice may need to be strengthened.

Dimensions within the CLASS Instructional Support Domain and Indicators within Dimensions		
Dimension = Concept development	**Dimension = Quality of Feedback**	**Dimension= Language Modeling**
Analysis and Reasoning	*Scaffolding*	*Frequent Conversation*
Creating	*Feedback Loops*	*Open-ended Questions*
Integration	*Prompting Thought Processes*	*Repetition and Extension*
Connections to the Real World	*Providing Information*	*Self- and Parallel Talk*
	Encouragement and Affirmation	*Advanced Language*

The study team looked at scores on the fourteen indicators within Instructional Support for all PEG teachers and compared average scores on those indicators for teachers who received relatively Low, Moderate, or High scores on the overall Instructional Support domain (Table 1 and Figure 11). The bar graph in Figure 16 clearly shows that the teachers with higher overall Instructional Support scores also had higher scores on each of the fourteen indicators compared to the rest of the teachers. However, it is also true that all teachers, including those with high overall ratings, struggled with some of the indicators, including Creating and Connections to the Real World (both in the Concept Development dimension) and Self- and Parallel Talk and Advanced Language (both in the Language Modeling dimension). The CLASS manual provides descriptions of what high levels of these indicators would look like. For example, for Self- and Parallel Talk, the manual gives examples of teachers who consistently map their own actions (self-talk) and the students' actions through language and description (parallel talk), as a way to expand the students' language.

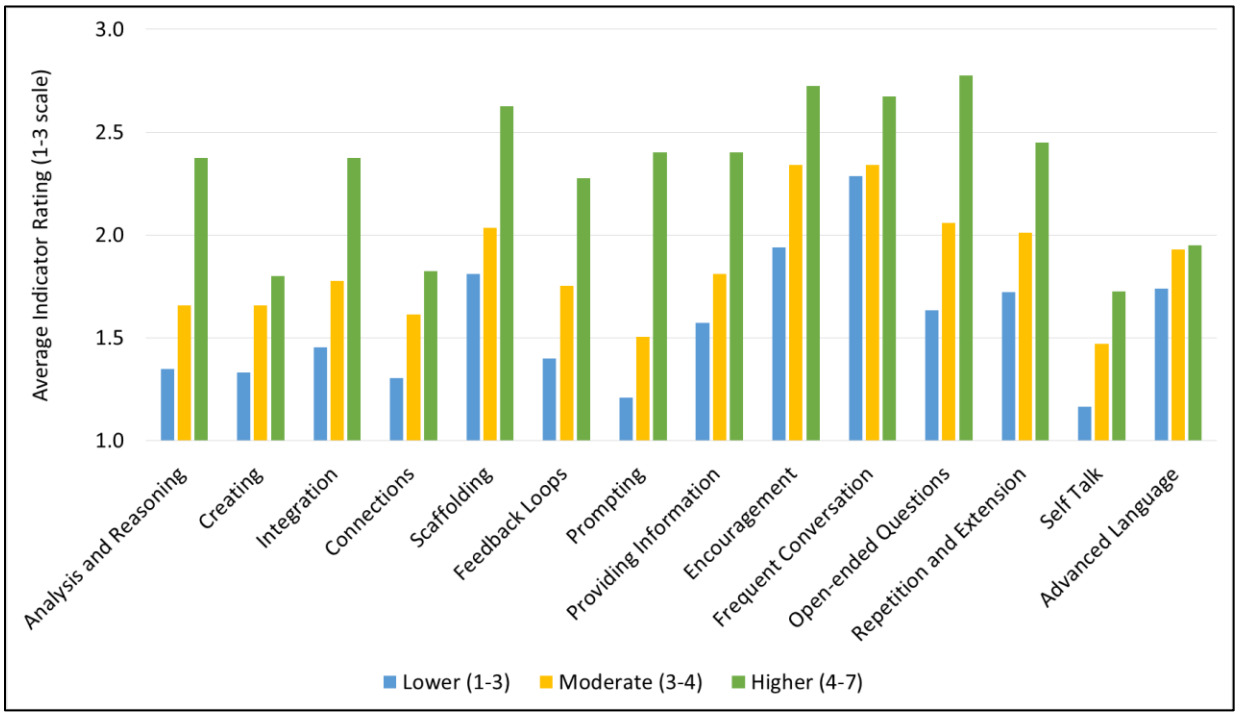
Table 1. Average Scores on Indicators within the CLASS Instructional Support Domain, by Overall Instructional Support Score

Indicators within the *Instructional Support* Domain [a]	PEG Classroom Scores on CLASS Instructional Support			
	Lower Scores (1 – 3) (n= 23)	Moderate Scores (3 - 4) (n= 17)	Higher Scores (4-7) (n= 8)	All Classrooms (n= 48)
Dimension = Concept Development				
Analysis and Reasoning	1.35	1.66	2.38	1.63
Creating	1.33	1.66	1.80	1.53
Integration	1.45	1.78	2.38	1.72
Connections to the Real World	1.30	1.61	1.83	1.50
Dimension = Quality of Feedback				
Scaffolding	1.81	2.04	2.63	2.03
Feedback Loops	1.40	1.75	2.28	1.67
Prompting Thought Processes	1.21	1.51	2.40	1.51
Providing Information	1.57	1.81	2.40	1.80
Encouragement and Affirmation	1.94	2.34	2.73	2.21
Dimension = Language Modeling				
Frequent Conversation	2.29	2.34	2.68	2.37
Open-ended Questions	1.63	2.06	2.78	1.98
Repetition and Extension	1.72	2.01	2.45	1.95
Self- and Parallel Talk	1.17	1.47	1.73	1.37
Advanced Language	1.74	1.93	1.95	1.84

[a] Indicators were scored on a 1-3 scale, with 1 = low, 2 = moderate, and 3 = high.

Notes. Numbers in this table are average indicator scores. For example, the first row of the table shows that teachers whose Instructional Support scores were "lower" had lower scores on the Analysis and Reasoning indicator (1-3 scale) compared to teachers whose Instructional Support scores were "moderate" or" higher" (1.35 out of 3 as compared to 1.66 or 2.38 out of 3).

Figure 11. Average Scores on Instructional Support Domain Indicators by Total Instructional Support Score (Lower/Moderate/Higher)



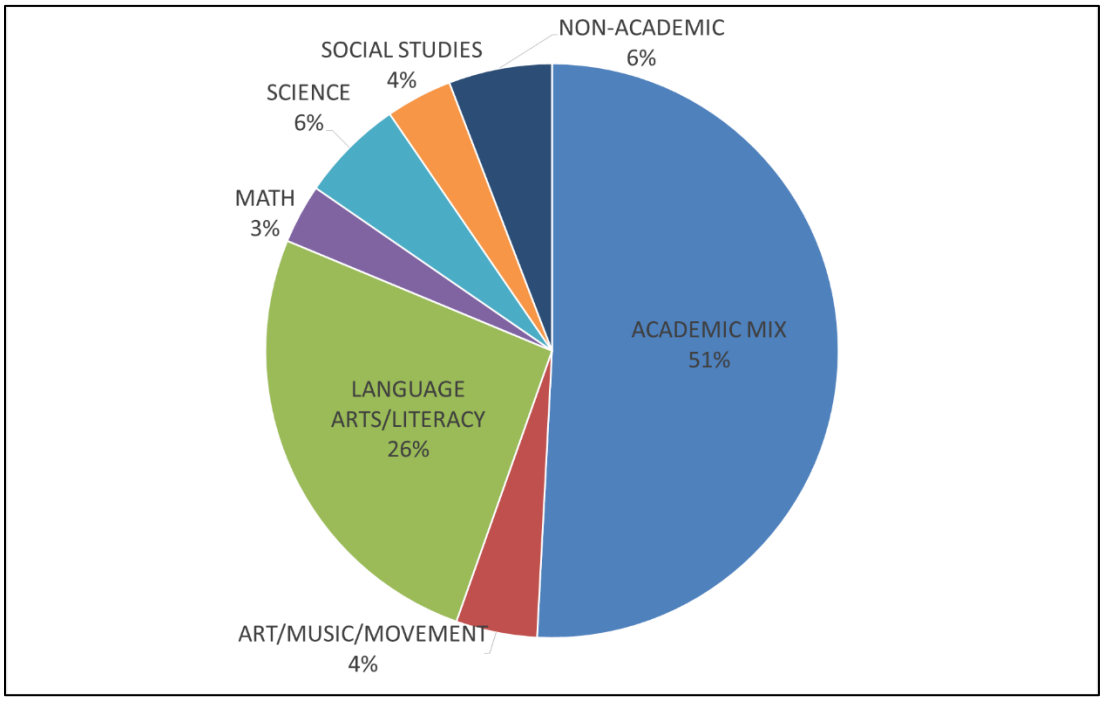
Relationship of Instructional Support Scores to the Content and Format of the Classroom during the Observation

During the CLASS observations this spring, for each twenty-minute coding cycle, observers also indicated: (a) the subject matter(s) or *content* that was the focus of the class and (b) the class *format* or grouping(s). For both, observers were instructed to select the code that represented what was happening during the majority of the cycle. The codes for content and format are shown in the table below.

Codes for Class Content	Codes for Class Format
Literacy/Language Arts	*Routine*
Math	*Meals/Snacks*
Science	*Whole Group*
Art	*Small Group*
Social Studies	*Individual time*
Other	*Free choice/centers*

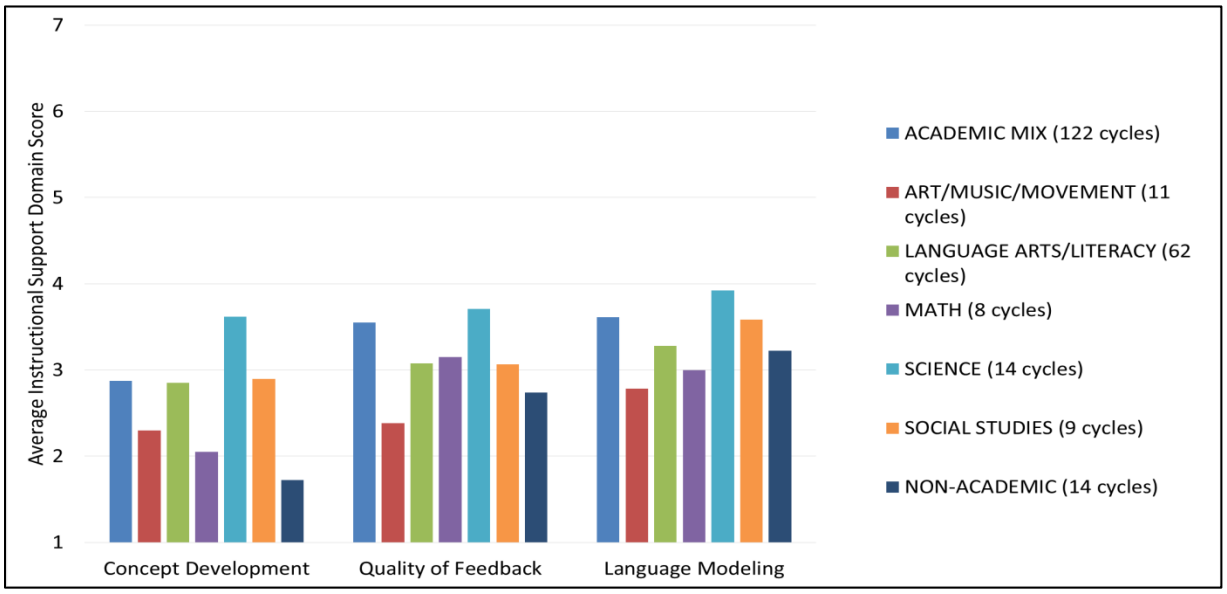
For Content, in half of the observation cycles, the classes were involved in a mix of various academic Content—Literacy, Math, Science, Social Studies, Art/Music/Movement – without a single content as the majority focus (122 or 51% of cycles), and in one-quarter of the observation cycles, the classes were involved in Language Arts/Literacy (26%) as the majority focus. Figure 12 shows the percentage of total cycles observed (five per classroom) by their majority content area.

Figure 12. All Observation Cycles by Majority Content



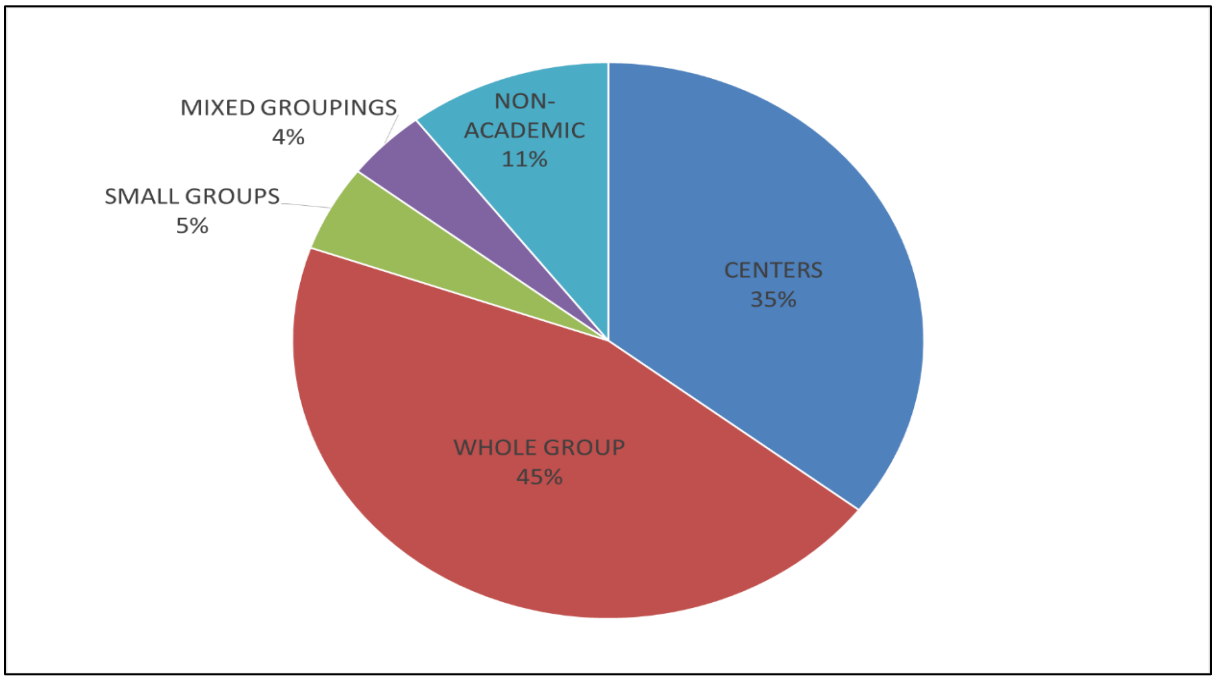
Average scores for Instructional Support were highest when Science alone was the major content of the observed cycle, followed closely by mixed academic content (see Exhibit 13). Average scores for Instructional Support were lowest when the content was Art/Music/Movement lessons and Non-academic content.

Exhibit 13. Average Observation Cycle Scores for CLASS Instructional Support by Majority Content



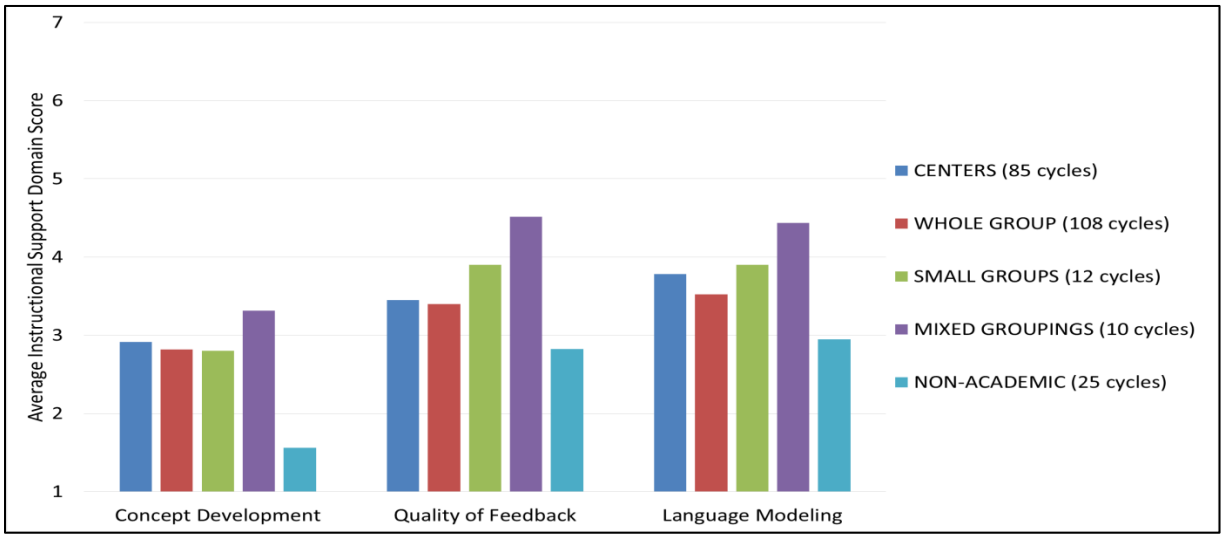
For Format, in almost half of the observation cycles, children were in Whole Group activities the majority of the time. They were primarily in Centers in about one-third of the observations. Exhibit 14 shows the percentage of total cycles observed (five per classroom) by their majority format.

Exhibit 14. All Observation Cycles by Majority Format



Average scores for Instructional Support were highest when children were in Mixed Groupings (where the teacher used multiple types of groupings within the same cycle), followed closely by Center groupings and Small Groups (see Figure 15). Average scores were lowest when children were in Whole Group activities and during Routines (non-academic activities like transitions or meals).[1]

Figure 15. Average Observation Cycle Scores for CLASS Instructional Support by Majority Format (Grouping)



[1] If the majority *content* of a cycle was coded as non-academic, it did not necessarily mean that the majority *format* of that cycle was non-academic (and vice-versa). For example, one cycle had a majority content of Math but a majority format of Non-Academic because most of the observation cycle was during a routine transition in the classroom (that happened to include math instruction).