

M A S S A C H U S E T T S

PERFORMANCE ASSESSMENT *for* LEADERS

# Annual Technical Report

## 2016–2017 Program Year



# Contents

Introduction .....	1
<b>PAL Assessment Summary</b> .....	<b>1</b>
<b>Pathways to Principal Licensure</b> .....	<b>2</b>
<b>PAL Assessment Development Summary</b> .....	<b>4</b>
<b>PAL Assessment Design</b> .....	<b>4</b>
Task 1: Leadership through a Vision for High Student Achievement .....	4
Task 2: Instructional Leadership for a Professional Learning Culture .....	4
Task 3: Leadership in Observing, Assessing, and Supporting Individual Teacher Effectiveness.....	4
Task 4: Leadership for Family Engagement and Community Involvement .....	4
<b>PAL Content Validity, Bias and Sensitivity, Pilot Study and Field Trial</b> .....	<b>5</b>
<b>PAL Operational Year One (2015–2016)</b> .....	<b>6</b>
<b>The PAL Scoring System</b> .....	<b>7</b>
<b>Indicator Scores</b> .....	<b>7</b>
<b>Rubric Scores and Descriptors</b> .....	<b>7</b>
<b>Overall Task Scores and Status</b> .....	<b>7</b>
<b>PAL Summary Score</b> .....	<b>7</b>
<b>Condition Codes and Incomplete Tasks</b> .....	<b>8</b>
<b>Scoring Model</b> .....	<b>8</b>
• Resolution .....	8
• Adjudication .....	8
<b>Scorer Recruitment and Training</b> .....	<b>8</b>
Scorer Training–Experienced Scorers. ....	8
Scorer Training–New Scorers.....	8
<b>The Scoring Process</b> .....	<b>9</b>
Online Distributed Scoring.....	9
Scorer Monitoring.....	9
<b>Operational Administration Year 2016–2017</b> .....	<b>10</b>
<b>Candidate Performance Summary</b> .....	<b>10</b>
<b>Scoring Agreement</b> .....	<b>28</b>
<b>References</b> .....	<b>30</b>

# Introduction

This report presents the results for the second Program Year (2016–2017) of the Massachusetts Performance Assessment for Leaders (PAL).

---

## PAL Assessment Summary

The PAL provides a measure of leadership candidates' readiness for initial school leader positions, informing licensure decisions while also supporting candidate learning and preparation program improvement. The assessment consists of four field-based performance tasks that allow candidates to demonstrate their leadership knowledge and skills in planning for an area of school improvement, facilitating a professional learning group, observing and giving feedback to a teacher, and engaging families and the community in improving student learning. Candidates produce written memos, reports, and video products as evidence of their accomplishment of each task.

PAL, which is aligned with state and national leadership standards and indicators, was developed with input from K–12 school and district leaders and higher education faculty. More information on the development and policy around implementation of the PAL is available on the [ESE web site](#).

PAL is designed as a summative assessment of a candidate's key leadership knowledge and skills. PAL consists of four performance assessment tasks of leadership knowledge and skills. The tasks ask licensure candidates to set direction by developing a plan for an area of school improvement, creating a professional learning culture among school staff, supporting individual teacher development through observation and feedback, and engaging families and community in improving student learning. Specifically, the four tasks comprise the following:

- Task 1: Leadership through a vision for high student achievement.
- Task 2: Instructional leadership for a professional learning culture.
- Task 3: Leadership in observing, assessing, and supporting individual teacher effectiveness.
- Task 4: Leadership for family engagement and community involvement.

Effective July 1, 2016, the Massachusetts Department of Elementary and Secondary Education awarded the contract for PAL administration to the Evaluation Systems group of Pearson. Pearson implemented task-based registration, scoring, and reporting. Candidates register for each task individually, for new or any retake submissions ([candidate and program website](#)).

In a [memo](#) dated 12/17/2015, the Assistant Director of Educator Effectiveness announced:

*Based on the recommendations of the PAL Technical Advisory and Standards Setting committees, the Commissioner of Elementary and Secondary Education established the performance assessment standards for PAL, resulting in two components of scoring that a candidate must meet or exceed:*

- *The minimum threshold score (cut score) for each task of 2.1 (on a 4-point scale), and*

## Massachusetts PAL 2016–2017 Technical Report

- *A total passing score for Program Year 2015–16, of 2.5 (on a 4-point scale). Beginning with Program Year 2016–17, the total passing score will shift to 2.75. The 2.5 total passing score is used in 2015–16 only, in order to support the first year of full implementation for PAL and efforts made by the field and leadership candidates to adapt to this new assessment.*

**Completers and Candidate Totals.** Based on the assessment model, there are two sets of candidate numbers: PAL Completers (candidates who have completed all four tasks in the 2016–2017 program year), and all PAL Candidates (candidates attempting at least one task in the 2016–2017 program year).

A summary of candidate numbers is as follows:

1. The N of PAL Completers (2016–2017) is 151; i.e., 151 candidates completed their final task in the 2016–2017 year, thereby obtaining scores on all four tasks within the PAL Assessment.
2. The N of all PAL Candidates (2016–2017) attempting at least one task is 323, as follows:

**Table 1. Number of Tasks Completed by PAL Candidates**

N of Tasks Completed	N of Candidates
4	151
3	49
2	51
1	72
<b>Total</b>	323

---

## Pathways to Principal Licensure

There are three pathways to principal licensure in Massachusetts, which are described immediately below: 1) completion of a state-approved preparation program, 2) an administrative apprenticeship/internship pathway, and 3) a panel review process.

**State-Approved Preparation Programs.** Such programs may be offered by public and private higher education institutions, districts, collaboratives, and non-profit organizations. Regardless of the type of organization, all Massachusetts providers are required to meet the same rigorous expectations for approval and undergo the same [processes associated with reviews](#).

**Administrative Apprenticeship/Internship.** This pathway, launched in 2001, was designed to enable districts to support aspiring education leaders by providing seminars and other leadership development learning experiences. Since 2012, candidates in this pathway have been required to complete at least 500 hours of internship experience and demonstrate proficiency in the [Professional Standards and Indicators for Administrative Leadership](#).

**Panel Review.** The panel review option is available to applicants who have completed an accredited leadership or management program and have had at least three years of administrative, leadership, or

***Massachusetts PAL 2016–2017 Technical Report***

management experience. Candidates seeking licensure through this option must compile information on their professional education and professional experience and be interviewed by a panel of experienced administrators and educators.

# PAL Assessment Development Summary

The Department of Elementary and Secondary Education began development of PAL in 2012 in partnership with key stakeholder groups after regulations passed in 2011 requiring that candidates must demonstrate that they are meeting the [Professional Standards for Administrative Leadership](#) by completing a performance assessment for initial license (603.CMR 7.10). Between 2012 and 2015, the Department worked with their selected development vendor (Bank Street College of Education) to create, pilot, and field test the PAL tasks. Effective September 1, 2014, the Commonwealth of Massachusetts required all applicants for principal licensure to complete four PAL assessment tasks. Applicants included individuals enrolled in leadership preparation programs and those pursuing licensure through the administrative apprenticeship/internship or panel review routes. Fall 2015 was the first operational year of the PAL.

---

## PAL Assessment Design

As summarized [here](#), the tasks of the PAL Assessment are as follows:

### **Task 1: Leadership through a Vision for High Student Achievement**

Focusing on the two pillars of highly effective schools, the instructional program (curriculum, instruction, and assessment) and school culture, a candidate will develop a school vision and improvement plan for one school-based priority area. The candidate will collect and analyze quantitative and qualitative data on student performance, student and teacher relationships and school culture, select a priority area for focus, document existing school programs, services, and practices, and develop a set of goals, objectives, and action strategies with input from school leaders and key stakeholder groups.

### **Task 2: Instructional Leadership for a Professional Learning Culture**

A candidate will demonstrate their capacity to foster a professional learning culture to improve student learning, by working with a small group of teachers using structured learning activities to improve the teachers' knowledge and skills. The candidate will support teachers in improving an existing curriculum, instruction, or assessment strategy, while documenting the process, teachers' team work, and improved practices.

### **Task 3: Leadership in Observing, Assessing, and Supporting Individual Teacher Effectiveness**

A candidate will demonstrate instructional leadership skills to plan for a teacher observation, observe, analyze the observation and student data, provide feedback, and plan support for an individual teacher. A candidate will document his or her work in the observation cycle and the quality and use of the feedback provided to teachers.

### **Task 4: Leadership for Family Engagement and Community Involvement**

A candidate will develop a proposal and implement one component to improve family engagement and community involvement in a school's priority area that is related to student

achievement or student health, recreation, or social needs that impact their learning. A candidate will work collaboratively with a work group representing school leadership, staff, families and community members, and students to select a priority area based on evidence of student needs, gather information related to family engagement and community involvement needs, develop a proposal, and implement one component with work group support.

---

## **PAL Content Validity, Bias and Sensitivity, Pilot Study and Field Trial**

The [PAL Field Trial Technical Report](#) documents the development process leading up to the first operational program year. As outlined in the report, the PAL assessment system was developed and refined through a standards-based design process to ensure content validity and alignment to the state standards and expectations for beginning school leaders.

Representatives from a number of Massachusetts preparation programs and pathways, as well as K–12 education leaders, worked together to draft the PAL Field Trial Technical Report. These content area experts served on either a design committee or a content validity committee. Members of each committee reviewed the four draft tasks and the assessment system before they were piloted to determine their importance and relevance in relation to 1) state and national leadership standards, 2) the research literature on effective school leadership, and 3) the committee members’ knowledge of the job of new leaders. Determining the content validity required addressing the question: “How well does the content of PAL represent core domains of school leadership knowledge and skills?” The two committees conducted follow-up reviews after the Pilot Study, made revisions to the PAL assessment before the Field Trial was launched in September 2014, revised again after the Field Trial, and revised before the Program Year 2015–16 was launched.

Additionally, a Bias Review Committee (composed of nine experienced educational leaders and program faculty with expertise in detecting varied forms of bias) was formed and provided input on the tasks, which were then revised to reduce potential bias and increase sensitivity.

Conclusions drawn from 1) the three content validation steps of Standards Alignment, Design Committee Validation, and Content Committee Validation, and 2) the two face validity activities of Pilot Study and Field Trial Study Candidate and Faculty Face Validation, were as follows:

“The PAL tasks have very good content validity, based on the strong agreement from the PAL design and content validity committees, and reinforced by the face validation from Pilot Study and Field Trial surveys of program faculty and candidates. Both committees strongly agreed that the four PAL tasks are aligned to the Massachusetts Standards for Administrative Leadership, provide authentic job-related experiences, and are relevant to the work that successful school leaders must be able to do. The strong agreement among the content validity committee members for all indicators and tasks exceeds professional standards for content validity (Wilson et al., 2012). These results were further confirmed by strongly positive agreement ratings for face validity in the Pilot Study and Field Trial surveys of preparation program faculty, and the positive agreement among most Pilot Study and the majority of the Field Trial candidates.” (Orr et al., 2016).

## **PAL Operational Year One (2015–2016)**

The [2015–2016 PAL Technical Report](#) documents the first operational program year of the PAL (2015–2016) including PAL completer and leadership preparation program feedback, as well as the assessment results for the first program year. As concluded in this report, “PAL task assessment evidence continues to support the use of PAL assessments as a requirement to determine candidate readiness for initial school leader licensure. Candidates’ experiences demonstrate that the tasks meet the performance assessment purposes of being authentic, challenging, educative, and relevant to their school settings. The assessment requirements are found to be well aligned to preparation programs and lead to modest curriculum and field work changes, with little negative consequence.” (Orr et al., 2016).

# The PAL Scoring System

Beginning with the 2016–2017 operational program year, all task submissions were double scored. As such, rules were established around double scoring, resolution, and reporting as outlined in this section of the report.

---

## Indicator Scores

Each PAL submission is reviewed by two independent scorers. The Indicator Scores are the average of the scores provided by the two scorers, including any applicable resolution. Scored indicators receive a numeric score between 1 and 4.

---

## Rubric Scores and Descriptors

Each Rubric Score is the average of its collection of indicator scores. Note that the average value reported is truncated and not rounded. Scored rubrics receive a numeric score between 1 and 4, with a descriptor indicating the level of attainment for that rubric, as follows:

Rubric Score Range	Descriptor
1.00 to 2.09	Beginning
2.10 to 2.74	Developing
2.75 to 3.49	Meeting
3.50 to 4.00	Exceeding

---

## Overall Task Scores and Status

The Overall Task Score is the truncated (unrounded) average of all the Rubric Scores within that task. Scored tasks receive a numeric score between 1 and 4. All tasks must meet or exceed a threshold score of 2.1. In addition to a numeric score, the Task Score Summary Report indicates whether or not the Overall Task Score met or exceeded the minimum threshold score of 2.1.

---

## PAL Summary Score

The PAL Summary Score is the average of the Overall Task Score values (best attempts) of all four submitted and scored tasks. Note that the average value reported is truncated and not rounded. In order to pass the PAL assessment, each task must meet the minimum threshold score of 2.1, and the PAL Summary Score must meet or exceed the PAL passing score of 2.75 (effective the 2016–2017 program year).

## Condition Codes and Incomplete Tasks

Any indicator score assigned a letter (e.g., “A”) instead of a numeric score is an indication that the submission or portions of the submission are deemed unscorable in accordance with the [PAL Submission Requirements](#). If a condition code is received for any indicator, the task in which that Condition Code was assigned would **not** be included in the PAL Summary Score. Any task that contains an indicator with a Condition Code will be unscorable and reported overall as “Incomplete.”

---

## Scoring Model

The following bullets summarize the scoring model applied starting with the 2016–2017 program year:

- Scorers evaluate the entire submitted task and apply scores by indicator.
  - All task submissions are double scored (that is, scored by two scorers independently).
  - Rater agreement is calculated by indicator, and evaluated through exact and adjacent scores.
  - Double-scored task submissions are evaluated by a scoring supervisor in the event a resolution or adjudication is required:
    - **Resolution:** If Scorer 1 and Scorer 2 are discrepant (i.e., more than 1 score point apart) on any indicator, the task is resolved by a scoring supervisor.
    - **Adjudication:** If Scorer 1 and Scorer 2 are on opposite sides of the task threshold score (2.1), the task is adjudicated by a scoring supervisor who scores the entire task submission.
  - If a portfolio does not need resolution or adjudication, then the average of Scorer 1 and Scorer 2 is reported to the candidate.
- 

## Scorer Recruitment and Training

PAL scoring for the 2016–2017 program year was conducted by a mix of experienced scorers trained initially by the previous vendor, Bank Street College of Education, and new scorers trained by Pearson. Scoring began immediately following the first submission deadline in the 2016–2017 program year. Scorer qualifications did not change from the prior program year.

**Scorer Training—Experienced Scorers.** Experienced scorers completed the scoring for the first two submission deadlines in the 2016–2017 program year. Training consisted of webinars conducted by supervisors, who led a group of scorers for each task through a discussion of exemplars and practice portfolios, followed by independent qualification through meeting the passing standard on a calibration portfolio. Each scorer discussed the results of the calibration with a supervisor after completing it, reviewing the rationale for all rubric scores even when the scorer had accurate scores.

**Scorer Training—New Scorers.** In order to become an official PAL scorer, educators must successfully complete scorer training and meet qualification standards. Training for scorers included both on-site instruction and further individualized online practice and discussion, totaling about 15 hours.

Experienced scorers joined new scorers during on-site training sessions conducted by Pearson in 2017. The on-site training included an orientation of scorers to the task, rubrics, and scoring system, and provides numerous opportunities to identify and evaluate evidence for each rubric.

After guided scoring through an exemplar portfolio, scorers independently scored sample PAL portfolios pre-selected by scoring supervisors, and then reviewed evidence and score justifications with the group. Scorers were then required to complete the independent scoring of an additional practice portfolio and review of the scores one-on-one with a supervisor. Scorers then scored a calibration portfolio within passing standards before becoming fully qualified to score.

---

## **The Scoring Process**

**Online Distributed Scoring.** Scoring is conducted by qualified scorers using an online distributed scoring system. Scorers are able to access task submissions through the secure online system, and are provided with training and support information for the online system.

**Scorer Monitoring.** Scorers are monitored through the use of multiple reports that provide information at the task and rubric indicator level on inter-rater reliability (exact agreement, adjacency, discrepancy rates), rate of scoring (total number and average time taken to score each portfolio), and backreading performance.

Scorers are systematically monitored by their supervisors through a backreading process that ensures they are applying scores accurately and consistently. Backreading is defined as supervisors scoring a previously scored portfolio for the purpose of reviewing the original scoring and providing feedback to the scorer. During backreading, a scoring supervisor applies scores and identifies key evidence to support the scores. After applying scores, supervisors review scores from the original scoring and review backreading scores with feedback to the original scorer as appropriate.

# Operational Administration Year 2016–2017

Following presents information on performance during the 2016–2017 program year. **Note that data is suppressed for any groups with fewer than 10 candidates (represented as “- -”).**

---

## Candidate Performance Summary

During Program Year 2016–17, 323 candidates submitted to Evaluation Systems at least one scorable PAL portfolio task toward the assessment for licensure (as shown in Table 2). This included 151 candidates who had completed all four tasks, 49 candidates who completed three tasks, 51 candidates who completed two tasks, and 72 candidates who completed one task. In contrast, 153 candidates completed all four tasks in Program Year 2015–16, and 422 completed all four tasks as part of the Field Trial.

**Table 2. Number of Candidates Who Submitted At Least One Task in Program Year 2016–17 for PAL by Cumulative Number of Tasks Completed**

Completed Tasks Submitted	Number of Candidates
4 tasks	151
3 tasks	49
2 tasks	51
1 task	72
<b>Total</b>	<b>323</b>

Note: This table excludes candidates with submissions being reviewed for irregularities and retakes.

One hundred and fifty-one (151) candidates completed PAL in 2016–17, including 108 who participated in a preparation program and 43 who pursued an alternative pathway to licensure (referred to here as PAL completers). As during the Field Trial and the 2015–16 program year, PAL completers in the 2016–2017 program year were primarily female and enrolled in or had recently completed a leadership preparation program.

Table 3. Assessment Pass Rates for Tasks by Demographic Characteristic

	Assessment (Completers Only)		Task 1: Leadership through a Vision for High Student Achievement		Task 2: Instructional Leadership for a Professional Learning Culture		Task 3: Leadership in Observing, Assessing, and Supporting Individual Teacher Effectiveness		Task 4: Leadership for Family Engagement and Community Involvement	
	N	% Pass	N	% Pass	N	% Pass	N	% Pass	N	% Pass
<b>All Completers</b>	151	91%	254	99%	235	99%	219	99%	217	96%
<b>Preparation Pathway</b>										
<b>Preparation Program</b>	108	91%	179	99%	174	98%	163	99%	158	96%
<b>Alternative Pathway-Internship</b>	41	93%	68	100%	58	100%	52	100%	56	96%
<b>Alternative Pathway-Panel Review</b>	--	--	--	--	--	--	--	--	--	--
<b>Out of State</b>	--	--	--	--	--	--	--	--	--	--
<b>Gender</b>										
<b>Male</b>	41	93%	65	100%	63	98%	64	98%	61	93%
<b>Female</b>	110	91%	189	99%	171	99%	155	99%	155	97%
<b>Not Reported</b>	--	--	--	--	--	--	--	--	--	--
<b>Race/Ethnicity</b>										
<b>Native American</b>	0	--	0	--	0	--	0	--	0	--
<b>Asian</b>	--	--	--	--	--	--	--	--	--	--
<b>Black</b>	--	--	--	--	--	--	--	--	--	--
<b>White</b>	129	91%	211	99%	194	98%	184	99%	177	96%

**Massachusetts PAL 2016–2017 Technical Report**

<b>Hispanic</b>	--	--	--	--	10	100%	--	--	10	90%
<b>Native American/Latino</b>	0	--	0	--	0	--	0	--	0	--
<b>Asian/Latino</b>										
<b>Black/Latino</b>	0	--	0	--	0	--	0	--	0	--
<b>Pacific</b>	--	--	--	--	--	--	--	--	--	--
<b>Pacific/Latino</b>	0	--	0	--	0	--	0	--	0	--
<b>White/Latino</b>	--	--	--	--	--	--	--	--	--	--
<b>Multiracial</b>	0	--	0	--	0	--	0	--	0	--
<b>Not reported</b>	--	--	11	100%	10	100%	--	--	10	100%

Table 3 shows the percentage by demographic attribute of completers and all takers passing all or some of the PAL tasks. These data show that the percentage distributions differ somewhat by program pathway and gender. PAL candidates from alternative pathways passed at a slightly higher rate than candidates through traditional preparation programs, although the N difference reflects about three times more candidates in the traditional preparation program route. Similarly, while male completers had a slightly higher pass rate than female completers, the N difference also reflects about three times more candidates are female than male.

Significance testing was conducted on all group comparisons by task and by total score. Findings showed no significant difference in performance across pathway, gender, or race/ethnicity both for task score comparisons and total score comparisons.

Table 4 shows the percentage distribution of 2016–17 PAL completers by performance level for the four tasks by demographic characteristic. As shown here, only a small percentage of candidates did not achieve the threshold of 2.1 for each task, with Task 4 showing a relatively high percentage of candidates scoring between 2.10 and 2.74 as compared to this performance range in the other tasks. Similarly, only a small percentage had exemplary scores for any task, with the highest percentage occurring in Task 2 (5%).

Table 4. Percentage Distribution of 2016-17 PAL Completers by Task, Performance Level, and Demographic Characteristics

	Task 1					Task 2				
	N	1.00-2.09	2.10-2.74	2.75-3.49	3.50-4.00	N	1.00-2.09	2.10-2.74	2.75-3.49	3.50-4.00
<b>Completers Only</b>	151	1	16	83	1	151	1	6	88	5
<b>Preparation Pathway</b>										
<b>Preparation Program</b>	108	1	15	83	1	108	1	6	87	6
<b>Alternative Pathway-Internship</b>	41	0	15	85	0	41	0	7	90	2
<b>Alternative Pathway-Panel Review</b>	--	--	--	--	--	--	--	--	--	--
<b>Out of State</b>	--	--	--	--	--	--	--	--	--	--
<b>Gender</b>										
<b>Male</b>	41	0	12	88	0	41	0	5	90	5
<b>Female</b>	110	1	17	81	1	110	1	6	87	5
<b>Race/Ethnicity</b>										
<b>Native American</b>	0	--	--	--	--	0	--	--	--	--
<b>Asian</b>	--	--	--	--	--	--	--	--	--	--
<b>Black</b>	--	--	--	--	--	--	--	--	--	--
<b>White</b>	129	1	16	82	1	129	1	5	88	5
<b>Hispanic</b>	--	--	--	--	--	--	--	--	--	--
<b>Native American/Latino</b>	0	--	--	--	--	0	--	--	--	--
<b>Asian/Latino</b>	0	--	--	--	--	0	--	--	--	--
<b>Black/Latino</b>	0	--	--	--	--	0	--	--	--	--

**Massachusetts PAL 2016–2017 Technical Report**

<b>Pacific</b>	0	--	--	--	--	0	--	--	--	--
<b>Pacific/Latino</b>	0	--	--	--	--	0	--	--	--	--
<b>White/Latino</b>	--	--	--	--	--	--	--	--	--	--
<b>Multiracial</b>	--	--	--	--	--	--	--	--	--	--
<b>Multiracial/Latino</b>	0	--	--	--	--	0	--	--	--	--
<b>Not reported</b>	--	--	--	--	--	--	--	--	--	--

Table 4. Percentage Distribution of 2016-17 PAL Completers by Task, Performance Level, and Demographic Characteristics

	Task 3					Task 4				
	N	1.00-2.09	2.10-2.74	2.75-3.49	3.50-4.00	N	1.00-2.09	2.10-2.74	2.75-3.49	3.50-4.00
<b>Completers Only</b>	151	1	8	89	2	151	2	23	74	1
<b>Preparation Pathway</b>										
<b>Preparation Program</b>	108	1	5	93	2	108	2	19	77	2
<b>Alternative Pathway-Internship</b>	41	0	17	80	2	41	2	32	66	0
<b>Alternative Pathway-Panel Review</b>	--	--	--	--	--	--	--	--	--	--
<b>Out of State</b>	--	--	--	--	--	--	--	--	--	--
<b>Gender</b>										
<b>Male</b>	41	0	10	90	0	41	2	24	73	0
<b>Female</b>	110	1	7	89	3	110	2	23	74	2
<b>Race/Ethnicity</b>										
<b>NatAmer</b>	--	--	--	--	--	--	--	--	--	--
<b>Asian</b>	--	--	--	--	--	--	--	--	--	--
<b>Black</b>	--	--	--	--	--	--	--	--	--	--
<b>White</b>	129	1	8	89	2	129	2	22	74	2
<b>Hispanic</b>	--	--	--	--	--	--	--	--	--	--
<b>NatAmer/Latino</b>	--	--	--	--	--	--	--	--	--	--
<b>Asian/Latino</b>	--	--	--	--	--	--	--	--	--	--
<b>Black/Latino</b>	--	--	--	--	--	--	--	--	--	--

**Massachusetts PAL 2016–2017 Technical Report**

<b>Pacific</b>	--	--	--	--	--	--	--	--	--	--
<b>Pacific/Latino</b>	--	--	--	--	--	--	--	--	--	--
<b>White/Latino</b>	--	--	--	--	--	--	--	--	--	--
<b>Multiracial</b>	--	--	--	--	--	--	--	--	--	--
<b>Multiracial/Latino</b>	--	--	--	--	--	--	--	--	--	--
<b>Not reported</b>	--	--	--	--	--	--	--	--	--	--

## Massachusetts PAL 2016–2017 Technical Report

Next, we examined the mean scores for the tasks (Table 5), rubrics (Table 6), and indicators (Table 7). Like prior program results, these tables show that the candidates' task scores are fairly similar, as shown in Table 5. The mean task scores are slightly higher for Tasks 2 and 3, with Task 4 having a slightly lower mean performance and slightly larger standard deviation.

Table 6 shows that average rubric scores range from 2.82 up to 3.06. The highest performing rubric was 3c (Provide Feedback and Suggest Support) and the lowest performing rubric was 4c (Analyze Feedback from Participants and Assess Leadership Skills).

Table 7 shows that average indicator scores range from 2.73 to 3.18. The highest performing indicator was 3c2 (Rapport and teacher engagement) and the lowest performing indicator as 4c2 (Assessment of leadership skills and practices). These indicators match up with the highest and lowest performing rubrics.

**Table 5. Descriptive Statistics of Tasks for 2016–17 PAL Completers**

Task	Task Score				
	N	Mean	S.D.	Min	Max
<b>Task 1: Leadership through a Vision for High Student Achievement</b>	151	2.91	0.24	1.80	3.72
<b>Task 2: Instructional Leadership for a Professional Learning Culture</b>	151	3.01	0.25	1.75	3.83
<b>Task 3: Leadership in Observing, Assessing, and Supporting Individual Teacher Effectiveness</b>	151	3.01	0.23	1.87	3.64
<b>Task 4: Leadership for Family Engagement and Community Involvement</b>	151	2.87	0.32	1.30	3.52

Table 6. Descriptive Statistics of Rubrics by Task for 2016–17 PAL Completers

		Rubric Score				
		N	Mean	S.D.	Min	Max
Task	Rubric					
Task 1: Leadership through a Vision for High Student Achievement	Rubric 1a: Investigate and Prepare a Vision	151	2.94	0.27	1.33	3.83
	Rubric 1b: Design an Integrated Plan for Strategies to Develop and Implement Improvement in the Priority Academic Area	151	2.93	0.27	1.83	3.83
	Rubric 1c: Assess and Analyze Feedback from Participants	151	2.89	0.31	2.00	3.50
Task 2: Instructional Leadership for a Professional Learning Culture	Rubric 2a: Plan to Facilitate Group Learning	151	3.04	0.25	1.50	4.00
	Rubric 2b: Enact a Professional Learning Culture to Support Team Learning	151	2.99	0.31	2.00	4.00
	Rubric 2c: Assess Team Learning to Improve Ongoing Group Learning	151	3.00	0.29	1.75	3.75
Task 3: Leadership in Observing, Assessing, and Supporting Individual Teacher Effectiveness	Rubric 3a: Plan	151	3.06	0.32	2.00	4.00
	Rubric 3b: Conduct the Observation	151	3.00	0.25	1.50	3.75
	Rubric 3c: Provide Feedback and Suggest Support	151	3.06	0.28	2.00	4.00
	Rubric 3d: Assess: Analyze and Identify Implications	151	2.92	0.40	2.00	4.00
Task 4: Leadership for Family Engagement and Community Involvement	Rubric 4a: Plan to Promote Family and Community Involvement	151	2.87	0.35	1.16	3.50
	Rubric 4b: Implement an Engagement or Involvement Strategy	151	2.94	0.39	1.50	3.50
	Rubric 4c: Analyze Feedback from Participants and Assess Leadership Skills	151	2.82	0.39	1.25	3.75

Table 7. Descriptive Statistics of Indicators by Task for 2016–17 PAL Completers

			Indicator Score				
			N	Mean	S.D.	Min	Max
Task	Rubric	Indicator					
<b>Task 1: Leadership through a Vision for High Student Achievement</b>	<b>Rubric 1a: Investigate and Prepare a Vision</b>	<b>Indicator 1a1: Data collection</b>	151	2.94	0.29	1.00	3.50
		<b>Indicator 1a2: Data analysis and priority definition</b>	151	2.95	0.33	1.00	4.00
		<b>Indicator 1a3: Evaluation of existing policies, practices, and programs</b>	151	2.92	0.33	2.00	4.00
	<b>Rubric 1b: Design an Integrated Plan for Strategies to Develop and Implement Improvement in the Priority Academic Area</b>	<b>Indicator 1b1: Vision plan and focus</b>	151	2.92	0.31	2.00	4.00
		<b>Indicator 1b2: Solicitation of input from teachers and other stakeholders</b>	151	2.94	0.34	2.00	4.00
		<b>Indicator 1b3: Plan details</b>	151	2.91	0.34	1.50	4.00
	<b>Rubric 1c: Assess and Analyze Feedback from Participants</b>	<b>Indicator 1c1: Plan feedback</b>	151	2.98	0.35	2.00	4.00
		<b>Indicator 1c2: Assessment of leadership skills and practices</b>	151	2.80	0.40	1.00	3.50

Table 7. Descriptive Statistics of Indicators by Task for 2016–17 PAL Completers

			Indicator Score				
			N	Mean	S.D.	Min	Max
<b>Task 2: Instructional Leadership for a Professional Learning Culture</b>	<b>Rubric 2a: Plan to Facilitate Group Learning</b>	<b>Indicator 2a1: Group identification and group formation</b>	151	3.04	0.25	2.00	4.00
		<b>Indicator 2a2: Group learning plan</b>	151	3.05	0.30	1.00	4.00
	<b>Rubric 2b: Enact a Professional Learning Culture to Support Team Learning</b>	<b>Indicator 2b1: Group process</b>	151	3.03	0.30	2.00	4.00
		<b>Indicator 2b2: Group learning and work</b>	151	2.95	0.39	2.00	4.00
	<b>Rubric 2c: Assess Team Learning to Improve Ongoing Group Learning</b>	<b>Indicator 2c1: Assessment of group process and group work</b>	151	3.03	0.35	1.50	4.00
		<b>Indicator 2c2: Assessment of leadership skills and practices</b>	151	2.98	0.33	1.50	4.00

Table 7. Descriptive Statistics of Indicators by Task for 2016–17 PAL Completers

			Indicator Score				
			N	Mean	S.D.	Min	Max
<b>Task 3: Leadership in Observing, Assessing, and Supporting Individual Teacher Effectiveness</b>	<b>Rubric 3a: Plan</b>	<b>Indicator 3a1: Observation focus selection</b>	151	3.10	0.35	2.00	4.00
		<b>Indicator 3a2: Pre-observation meeting</b>	151	3.02	0.39	1.00	4.00
	<b>Rubric 3b: Conduct the Observation</b>	<b>Indicator 3b1: Use and application of teacher observation rubric</b>	151	3.00	0.33	1.00	4.00
		<b>Indicator 3b2: Description of observations</b>	151	2.99	0.23	2.00	4.00
	<b>Rubric 3c: Provide Feedback and Suggest Support</b>	<b>Indicator 3c1: Feedback content</b>	151	3.03	0.30	2.00	4.00
		<b>Indicator 3c2: Rapport and teacher engagement</b>	151	3.18	0.38	2.00	4.00
		<b>Indicator 3c3: Teacher development</b>	151	2.98	0.38	2.00	4.00
	<b>Rubric 3d: Assess: Analyze and Identify Implications</b>	<b>Indicator 3d1: Assessment of leadership skills and practices</b>	151	2.92	0.40	2.00	4.00

Table 7. Descriptive Statistics of Indicators by Task for 2016–17 PAL Completers

			Indicator Score				
			N	Mean	S.D.	Min	Max
<b>Task 4: Leadership for Family Engagement and Community Involvement</b>	<b>Rubric 4a: Plan to Promote Family and Community Involvement</b>	<b>Indicator 4a1: Investigation of the priority area</b>	151	2.84	0.42	1.00	3.50
		<b>Indicator 4a2: Investigation of work group engagement</b>	151	2.90	0.42	1.00	3.50
		<b>Indicator 4a3: Preparation of the plan, including strategies</b>	151	2.86	0.46	1.00	3.50
	<b>Rubric 4b: Implement an Engagement or Involvement Strategy</b>	<b>Indicator 4b1: Implementation of the strategy</b>	151	2.94	0.39	1.50	3.50
		<b>Rubric 4c: Analyze Feedback from Participants and Assess Leadership Skills</b>	<b>Indicator 4c1: Assessment and analysis of feedback on the family and community engagement plan and strategy</b>	151	2.91	0.51	1.00
	<b>Indicator 4c2: Assessment of leadership skills and practices</b>		151	2.73	0.44	1.50	4.00

## **Massachusetts PAL 2016–2017 Technical Report**

During Program Year 2016–2017, candidates were required to achieve a higher composite average score (2.75) to pass than they were during the previous program year (2.5). For this reason, the percentage of PAL completers who passed (and how well) was evaluated. The evaluation results show, in Table 8, that only 8 percent had scores that did not meet or exceeded the total average composite PAL score level (2.75) for 2016–17.

**Table 8. Percentage Distribution of 2016–17 PAL Completers by Total Average PAL Score**

<b>Score Range</b>	<b>Number</b>	<b>Percent</b>
<b>Less than 2.50</b>	3	2.0
<b>2.50–2.74</b>	9	6.0
<b>2.75–2.99</b>	77	51.0
<b>3.00 and above</b>	62	41.1
<b>All Completers</b>	151	100.0

The PAL completers' total PAL scores were compared by preparation pathway, gender, and race/ethnicity, and the results are shown in Table 9. The overall mean was 2.95 with a 0.18 standard deviation. Using t-tests for comparison of pairs by pathway and gender, the results were not statistically significant.

Table 9. Descriptive Statistics of Total Scores by Demographic Attributes of 2016–17 PAL Completers

	Total Score				
	N	Mean	S.D.	Min	Max
All Completers	151	2.95	0.18	1.90	3.38
Preparation Pathway					
Preparation Program	108	2.97	0.20	1.90	3.38
Alternative Pathway-Internship	41	2.91	0.14	2.51	3.18
Alternative Pathway-Panel Review	--	--	--	--	--
Out of State	--	--	--	--	--
Gender					
Male	41	2.94	0.13	2.58	3.31
Female	110	2.95	0.20	1.90	3.38
Race/Ethnicity					
Native American	0	--	--	--	--
Asian	--	--	--	--	--
Black	--	--	--	--	--
White	129	2.95	0.19	1.90	3.38
Hispanic	--	--	--	--	--
Native American/Latino	0	--	--	--	--
Asian/Latino	0	--	--	--	--
Black/Latino	0	--	--	--	--
Pacific	0	--	--	--	--
Pacific/Latino	0	--	--	--	--
White/Latino	--	--	--	--	--
Multiracial	--	--	--	--	--
Multiracial/Latino	0	--	--	--	--
Not reported	--	--	--	--	--

Next, we examined the influence of demographic attributes on candidate performance. While overall scores showed no statistically significant difference, there was a consistent pattern in the differences

**Massachusetts PAL 2016–2017 Technical Report**

between tasks by preparation pathway but not by gender. As shown in Table 10, PAL completers from preparation programs scored higher than PAL completers from alternative pathways in all tasks. There was not a consistent difference by gender; however, male and female PAL completers scored somewhat differently on the various tasks, with male PAL completers scoring higher on Task 2 and female PAL completers scoring higher on the other tasks.

Table 10. Descriptive Statistics of Task Scores by Demographic Attributes of 2016–17 PAL Completers

	Task 1					Task 2					Task 3					Task 4				
	N	Mean	S.D.	Min	Max	N	Mean	S.D.	Min	Max	N	Mean	S.D.	Min	Max	N	Mean	S.D.	Min	Max
<b>Completers Only</b>	151	2.91	0.24	1.80	3.72	151	3.01	0.25	1.75	3.83	151	3.01	0.23	1.87	3.64	151	2.87	0.32	1.30	3.52
<b>Preparation Pathway</b>																				
<b>Preparation Program</b>	108	2.93	0.25	1.80	3.72	108	3.02	0.27	1.75	3.83	108	3.02	0.22	1.87	3.64	108	2.90	0.34	1.30	3.52
<b>Alternative Pathway-Internship</b>	41	2.89	0.21	2.11	3.38	41	2.99	0.20	2.41	3.50	41	2.98	0.26	2.41	3.50	41	2.80	0.26	2.00	3.30
<b>Alternative Pathway-Panel Review</b>	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
<b>Out of State</b>	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
<b>Gender</b>																				
<b>Male</b>	41	2.91	0.18	2.50	3.44	41	3.03	0.20	2.66	3.75	41	2.98	0.20	2.54	3.41	41	2.85	0.29	2.00	3.41
<b>Female</b>	110	2.92	0.26	1.80	3.72	110	3.00	0.26	1.75	3.83	110	3.01	0.24	1.87	3.64	110	2.88	0.33	1.30	3.52
<b>Race/Ethnicity</b>																				
<b>Native American</b>	0	--	--	--	--	0	--	--	--	--	0	--	--	--	--	0	--	--	--	--
<b>Asian</b>	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
<b>Black</b>	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
<b>White</b>	129	2.91	0.25	1.80	3.72	129	3.01	0.25	1.75	3.83	129	3.00	0.23	1.87	3.64	129	2.88	0.33	1.30	3.52
<b>Hispanic</b>	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
<b>Native American/Latino</b>	0	--	--	--	--	0	--	--	--	--	0	--	--	--	--	0	--	--	--	--
<b>Asian/Latino</b>	0	--	--	--	--	0	--	--	--	--	0	--	--	--	--	0	--	--	--	--
<b>Black/Latino</b>	0	--	--	--	--	0	--	--	--	--	0	--	--	--	--	0	--	--	--	--

**Massachusetts PAL 2016–2017 Technical Report**

<b>Pacific</b>	0	--	--	--	--	0	--	--	--	--	0	--	--	--	--	0	--	--	--	--
<b>Pacific/Latino</b>	0	--	--	--	--	0	--	--	--	--	0	--	--	--	--	0	--	--	--	--
<b>White/Latino</b>	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
<b>Multiracial</b>	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
<b>Multiracial/Latino</b>	0	--	--	--	--	0	--	--	--	--	0	--	--	--	--	0	--	--	--	--
<b>Not reported</b>	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

As was done with the prior year’s results, task scores were correlated to evaluate the degree of association. Again, the four factors have a positive correlation for Task 1, 2, 3 and 4, with Task 1 and Task 4 holding the highest positive task-based correlations.

**Table 11. Factor Correlation for 2016–17 PAL Tasks**

Pearson Correlation Coefficients				
	Task 1	Task 2	Task 3	Task 4
Task 1	1.00000	0.26941	0.26673	0.42166
Task 2	0.26941	1.00000	0.21296	0.34311
Task 3	0.26673	0.21296	1.00000	0.38138
Task 4	0.42166	0.34311	0.38138	1.00000

Note: All correlation coefficients are significant at the  $p < .01$  level.  $N=151$ .

The PAL completer score results were positive for Program Year 2016–17, and consistent with prior program years, showing some improvement. Overall, PAL completers scored higher than prior years, and PAL completers from preparation programs appeared to have performed better on the four tasks overall than those who were not from preparation programs, while male and female candidates appeared to have performed similarly to each other.

The relatively small number of completers hindered statistical comparison of candidate performance and evaluation of the scores through factor analysis and reliability analyses. Nonetheless, correlations of candidate scores between tasks show that these tasks continue to be independent measures with modest degrees of association (the strongest being between tasks 1 and 4 and the weakest between tasks 1 and 3).

---

## Scoring Agreement

Scoring agreement was determined using submissions that were scored by two scorers, and the results were used to estimate scoring reliability. Exact agreement rates (scorers assigning the same exact score) and total agreement rates (scorers assigning either the same or adjacent scores) were calculated for each indicator.

Table 12 presents the results of the rater agreement and Kappa-N. The Kappa-N provides chance-corrected total agreement, or inter-rater agreement measures, that result from removing total agreement that may have occurred randomly (Brennan & Prediger, 1981). Chance-corrected agreement ranges from zero to one, with higher values representing greater levels of agreement. The table below shows kappa n ranged from 0.89 (indicators 4b.1 and 4c.2) to 1.00 (indicators 2b.1, 3a.1, and 3c.3). These outcomes support the conclusion that scorers tend to assign scores within +/- 1 (exact or adjacent) and rarely assign scores that differ by more than 1 point (discrepant).

Exact rates, indicating the percentage of cases where scorers scoring the same portfolio assigned the same score, are above 50 percent on all indicators. Total agreement rates, as shown in Table 12, indicate

that independent scorers scoring the same submission assigned the same or adjacent scores over 96 percent of the time for all indicators.

**Table 12. Rater Agreement and Inter-rater Reliability by Indicator**

Indicator	N of Items Scored	Rater Agreement			Kappa N
		Percent Agreement	Percent Exact	Percent Adjacent	
1a.1	267	1.00	0.78	0.21	0.99
1a.2	267	1.00	0.75	0.25	0.99
1a.3	267	0.99	0.75	0.24	0.97
1b.1	267	1.00	0.73	0.27	0.99
1b.2	267	0.98	0.65	0.33	0.94
1b.3	267	0.99	0.67	0.32	0.97
1c.1	267	0.99	0.73	0.25	0.97
1c.2	267	0.99	0.64	0.35	0.98
2a.1	243	1.00	0.86	0.14	0.99
2a.2	243	0.99	0.82	0.16	0.97
2b.1	243	1.00	0.83	0.17	1.00
2b.2	243	0.99	0.70	0.29	0.98
2c.1	243	1.00	0.76	0.23	0.99
2c.2	243	1.00	0.66	0.33	0.99
3a.1	229	1.00	0.79	0.21	1.00
3a.2	229	0.99	0.79	0.20	0.98
3b.1	229	0.98	0.75	0.23	0.94
3b.2	229	0.99	0.84	0.15	0.98
3c.1	229	1.00	0.81	0.19	0.99
3c.2	229	1.00	0.72	0.27	0.99
3c.3	229	1.00	0.81	0.19	1.00
3d.1	229	1.00	0.73	0.26	0.99
4a.1	236	0.96	0.67	0.29	0.90
4a.2	236	0.98	0.64	0.34	0.94
4a.3	236	0.97	0.65	0.32	0.93
4b.1	236	0.96	0.62	0.33	0.89
4c.1	236	0.97	0.59	0.37	0.91
4c.2	236	0.96	0.54	0.42	0.89

## References

- Brennan, R. L., & Prediger, D. J. (1981). Coefficient Kappa: Some uses, misuses, and alternatives. *Educational and Psychological Measurement*, 41(3), 687-699.
- Orr, M. T., Pecheone, R. L., Shear, B., Hollingworth, L., & Beaudin, B. (2016). Massachusetts Performance Assessment for Leaders (PAL) Technical Report: Summary of Validity and Reliability Studies for 2014-15 Field Trial of PAL. New York: Bank Street College of Education. Retrieved from <http://www.doe.mass.edu/pal/TechnicalReport.docx>
- Orr, M. T., Hollingworth, L., & Beaudin, B. (2016). Massachusetts Performance Assessment for Leaders (PAL) Technical Report Follow-Up: Summary of Feedback and Performance Studies for 2015-16 Program Year of PAL. New York: Bank Street College of Education. Retrieved from <http://www.doe.mass.edu/pal/1516TechnicalReport.docx>
- Wilson, F. R., Pan, W., & Schumsky, D. A. (2012). Recalculation of the critical values for Lawshe's Content Validity Ratio. *Measurement and Evaluation in Counseling and Development*, 45(3), 197-210.