

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

PERFORMANCE ASSESSMENT *for* LEADERS

Annual Technical Report

2018–2019 Program Year



Contents

Introduction	1
PAL Assessment Summary	1
Pathways to Principal Licensure	2
PAL Assessment Development Summary	3
PAL Assessment Design	3
Task 1: Leadership through a Vision for High Student Achievement	3
Task 2: Instructional Leadership for a Professional Learning Culture	3
Task 3: Leadership in Observing, Assessing, and Supporting Individual Teacher Effectiveness.....	3
Task 4: Leadership for Family Engagement and Community Involvement	4
PAL Content Validity, Bias and Sensitivity, Pilot Study, and Field Trial	4
The PAL Scoring System	6
Indicator Scores	6
Rubric Scores and Descriptors	6
Overall Task Scores and Status	6
PAL Summary Score	6
Condition Codes and Incomplete Tasks	7
Scoring Model	7
• Resolution	7
• Adjudication	7
Scorer Recruitment and Training	7
Scorer Training—Experienced Scorers.....	7
Scorer Training—New Scorers	7
The Scoring Process	8
Online Distributed Scoring.....	8
Scorer Monitoring.....	8
Operational Administration Year 2018–2019	9
Candidate Performance Summary	9
Scoring Agreement	27
References	30
Appendix A	31
TASK 1	31

Massachusetts PAL 2018–2019 Technical Report

TASK 2	33
TASK 3	35
TASK 4	37

Introduction

This report presents the results for the fourth program year (2018–2019) 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 can be found on the [DESE 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. Please see the [candidate and program website](#) for more information.

Pathways to Principal Licensure

There are three pathways to principal licensure in Massachusetts: 1) completion of a state-approved preparation program, 2) an administrative apprenticeship/internship pathway, and 3) a panel review process. All three require a candidate to pass the PAL assessment.

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 internship experience and demonstrate proficiency in the [Professional Standards 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 the required number of years of administrative, leadership, or 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' teamwork, 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

Massachusetts PAL 2018–2019 Technical Report

preparation program faculty, and the positive agreement among most Pilot Study and the majority of the Field Trial candidates.” (Orr et al., 2016).

The PAL Scoring System

The PAL assessment includes indicators grouped under rubrics that are combined to create an overall score for each task. 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. These rules were then applied in subsequent operational program years.

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. “Resolution” is the term for a description of scoring at the indicator level. “Adjudication” refers to resolving total task scores that fall on either side of the cut score. 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 as of the 2016–2017 program year).

Condition Codes and Incomplete Tasks

Any indicator score assigned a letter (e.g., “B”) 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.” Effective as of the 2018–19 program year, Condition Code A (Work is not blinded) is no longer applied.

Scoring Model

The following bullets summarize the scoring model applied for the 2017–2018 program year:

- Scorers evaluate the entire submitted task and apply scores by indicator.
- All task submissions are double scored (i.e., 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 2018–2019 program year was conducted by a mix of trained experienced scorers and new scorers. Scorer qualifications did not change from the prior program year.

Scorer Training—Experienced Scorers. Experienced scorers completed the scoring for the first submission deadline in the 2018–2019 program year. Training consisted of practice portfolio discussions with supervisors, followed by independent qualification through meeting the passing standard on two calibration portfolios. Each scorer discussed the results of the calibrations with a supervisor after completing them, 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 2019. The on-site training included an orientation of scorers to the task, rubrics, and scoring system, and provided 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 2018–2019

The following presents information on performance during the 2018–2019 program year.

Note that data is suppressed for any groups with fewer than 10 candidates (represented as “-”). Groups with no data are represented with an N of “0”.

Candidate Performance Summary

Completers and All Takers Totals. Based on the assessment model, there are two sets of candidate numbers for the 2018–2019 report.

PAL Completers: Candidates who have taken all four tasks and at least one scorable task was submitted during this Program Year (i.e., between June 15, 2018, and June 20, 2019). If a candidate submitted at least one scorable task during this period, the data includes the candidate’s best score on all attempts for tasks submitted between September 30, 2016, and June 20, 2019. Reports based on Completers’ data, therefore, include best attempts only.

All Takers: Candidates who have submitted at least one scorable task between June 15, 2018, and June 20, 2019.

A summary of candidate numbers is as follows:

1. The N of PAL Completers is 307; that is, 307 candidates completed their **final** task in the 2018–2019 program year, thereby obtaining scores on all four tasks within the PAL Assessment.
2. The N of All Takers (2018–2019) attempting at least one task is 498, as follows:

Table 1. Number of Tasks Completed by PAL Candidates

N of Tasks Completed	N of Candidates
4	212
3	45
2	80
1	161
Total	498

During Program Year 2018–19, 498 individual candidates submitted to Pearson at least one scorable PAL portfolio task toward the assessment for licensure (as shown in Table 1). This included 212 candidates who completed all four tasks, 45 candidates who completed three tasks, 80 candidates who completed two tasks, and 161 candidates who completed one task. In contrast, fewer (379) individual candidates submitted at least one scorable PAL portfolio task and fewer (147) candidates completed all four tasks in Program Year 2017–2018.

Massachusetts PAL 2018–2019 Technical Report

Table 2 shows the percentage by demographic attribute of completers and all takers passing all or some of the PAL tasks. The first column includes the pass rate for all candidates who finished (“completers”) the PAL assessment by submitting their **final** task and passing the PAL in the 2018–2019 program year. This includes individuals who began submitting tasks in prior (2016–2017 or 2017–2018) program years, but only passed upon submitting their final task in the 2018–2019 program year. The first column also includes candidates who submitted all four tasks and passed the PAL assessment during the 2018–2019 program year only. The overall pass rate for the PAL assessment during the 2018–2019 program year was 91% (N=307).

Additional note, as stated previously: For all tables, Data is suppressed for any groups with fewer than 10 candidates (represented as “-”). Groups with no data are represented with an N of “0”.

For Table 2, data for the Native American group is shown as 0 / 0% percent because there were no data for that group in these categories. Had there been an N<10, data columns would be shown as “-” to suppress data.

Table 2 data show that the percentage distributions differ somewhat by program pathway and gender. PAL candidates from completing internships as an alternative pathway passed at a slightly lower rate than candidates from traditional preparation programs, although the N difference reflects about three times more candidates in the traditional preparation program route. Over twice as many females completed the assessment as males, although both groups had similar pass rates. However, there are differences between pass rates for candidates submitting Tasks 1, 2, and 4 in the 2018–2019 program year, with the pass rate for females about 3% higher than the pass rate for males.

Table 2. Best Attempt Assessment Pass Rates by Demographic Characteristic for Candidates Submitting Tasks in 2018–2019

	Assessment (Completed Final Task During 2018–2019)		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
Candidates	307	91%	366	97%	323	98%	313	100%	302	97%
Preparation Pathway										
Preparation Program	216	93%	265	98%	229	99%	220	100%	215	97%
Alternative Pathway— Internship	78	90%	86	94%	81	95%	80	100%	75	99%
Alternative Pathway— Panel Review	12	75%	12	92%	11	100%	12	100%	11	91%
Out of State	--	--	--	--	--	--	--	--	--	--
Gender										
Male	88	91%	112	99%	100	100%	96	100%	89	96%
Female	218	92%	252	96%	222	97%	216	100%	212	98%
Not Reported	--	--	--	--	--	--	--	--	--	--
Race/Ethnicity										
Native American	0	0%	0	0%	0	0%	0	0%	0	0%
Asian	--	--	--	--	--	--	--	--	--	--
Black	13	92%	13	100%	15	100%	19	100%	16	100%
White	262	91%	314	96%	276	98%	260	100%	251	97%

Massachusetts PAL 2018–2019 Technical Report

	Assessment (Completed Final Task During 2018–2019)		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
Hispanic	--	--	--	--	--	--	--	--	--	--
Native American/Latino	0	0%	0	0%	0	0%	0	0%	0	0%
Asian/Latino	--	--	--	--	--	--	--	--	--	--
Black/Latino	--	--	--	--	--	--	--	--	--	--
Pacific	--	--	0	0%	0	0%	0	0%	--	--
Pacific/Latino	0	0%	0	0%	0	0%	0	0%	0	0%
White/Latino	--	--	--	--	--	--	--	--	--	--
Multiracial	--	--	--	--	--	--	--	--	--	--
Multiracial/Latino	--	--	--	--	--	--	--	--	--	--
Not reported	--	--	--	--	--	--	--	--	--	--

Massachusetts PAL 2018–2019 Technical Report

Significance testing was conducted on all group comparisons by task and by total score. Findings showed no significant difference in performance across gender, or race/ethnicity both for task score comparisons and total score comparisons. Some differences were found in performance across pathways, including task scores for Task 2 (All Takers), Task 1 out of state (Completers), and Total Score by pathway for All Takers and Completers.

Table 3 shows the percentage distribution by performance level and demographic characteristic of the best attempt by PAL candidates who submitted at least one scorable task in the 2018–2019 program year. As shown here, only a small percentage of candidates did not achieve the threshold of 2.1 for each task, with the majority of candidates scoring in the 2.75–3.49 range. Only a small percentage of candidates had exemplary scores for Tasks 1–3, and no candidates had exemplary scores for Task 4.

Table 3. Best Attempt Percentage Distribution by Task, Performance Level, and Demographics for All Takers Submitting Tasks in 2018–2019

	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
Candidates	366	3	14	82	1	323	2	15	80	2
Preparation Pathway										
Preparation Program	265	2	14	84	1	229	1	13	83	3
Alternative Pathway— Internship	86	6	14	79	1	81	5	22	73	0
Alternative Pathway— Panel Review	12	8	17	75	0	11	0	27	73	0
Out of State	--	--	--	--	--	--	--	--	--	--
Gender										
Male	112	1	18	79	2	100	0	16	81	3
Female	252	4	12	83	0	222	3	15	81	2
Not Reported	--	--	--	--	--	--	--	--	--	--
Race/Ethnicity										
Native American	0	--	--	--	--	0	--	--	--	--
Asian	6	0	0	100	0	6	0	0	83	17
Black	13	0	23	77	0	15	0	33	67	0
White	314	4	13	83	1	276	2	16	81	2
Hispanic	8	0	38	63	0	6	0	17	83	0
Native American/Latino	0	--	--	--	--	0	--	--	--	--
Asian/Latino	--	--	--	--	--	--	--	--	--	--

Massachusetts PAL 2018–2019 Technical Report

	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
Black/Latino	--	--	--	--	--	--	--	--	--	--
Pacific	--	--	--	--	--	0	--	--	--	--
Pacific/Latino	0	--	--	--	--	0	--	--	--	--
White/Latino	--	--	--	--	--	--	--	--	--	--
Multiracial	--	--	--	--	--	--	--	--	--	--
Multiracial/Latino	--	--	--	--	--	0	--	--	--	--
Not reported	--	--	--	--	--	--	--	--	--	--

	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
Candidates	313	0	14	85	1	302	3	26	69	2
Preparation Pathway										
Preparation Program	220	0	15	84	1	215	3	26	70	2
Alternative Pathway— Internship	80	0	9	91	0	75	1	29	67	3
Alternative Pathway— Panel Review	12	0	33	67	0	11	9	18	73	0
Out of State	--	--	--	--	--	--	--	--	--	--
Gender										
Male	96	0	13	85	2	89	4	22	71	2

Massachusetts PAL 2018–2019 Technical Report

	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
Female	216	0	14	86	0	212	2	27	69	2
Not Reported	--	--	--	--	--	--	--	--	--	--
Race/Ethnicity										
NatAmer	0	--	--	--	--	0	--	--	--	--
Asian	--	--	--	--	--	--	--	--	--	--
Black	19	0	26	74	0	16	0	31	63	6
White	260	0	13	86	1	251	3	25	70	2
Hispanic	--	--	--	--	--	--	--	--	--	--
NatAmer/Latino	0	--	--	--	--	0	--	--	--	--
Asian/Latino	0	--	--	--	--	0	--	--	--	--
Black/Latino	--	--	--	--	--	--	--	--	--	--
Pacific	0	--	--	--	--	--	--	--	--	--
Pacific/Latino	0	--	--	--	--	0	--	--	--	--
White/Latino	--	--	--	--	--	--	--	--	--	--
Multiracial	--	--	--	--	--	--	--	--	--	--
Multiracial/Latino	--	--	--	--	--	--	--	--	--	--
Not reported	--	--	--	--	--	--	--	--	--	--

Massachusetts PAL 2018–2019 Technical Report

Next, we examined the mean scores for the tasks (Table 4), rubrics (Table 5), and indicators (Table 6). Table 4 shows some differences in candidate scores across tasks. Mean task scores are higher for Tasks 1 and 3, with (as in prior program years) Task 4 having a lower mean performance and slightly larger standard deviation.

Table 5 shows that average rubric scores range from 2.81 up to 2.98. The highest performing rubric was 3a (Plan). The lowest performing rubric was 4c (Analyze Feedback from Participants and Assess Leadership Skills).

Table 6 shows that average indicator scores range from 2.79 to 3.08. The highest performing indicator was 3c2 (Rapport and teacher engagement) and the lowest performing indicator was 2b2 (Group learning and work).

Table 4. Descriptive Statistics of Tasks for 2018–19 All Takers

Task	Task Score				
	N	Mean	S.D.	Min	Max
Task 1: Leadership through a Vision for High Student Achievement	366	2.93	0.31	1.33	3.66
Task 2: Instructional Leadership for a Professional Learning Culture	323	2.92	0.29	1.66	3.83
Task 3: Leadership in Observing, Assessing, and Supporting Individual Teacher Effectiveness	313	2.94	0.21	2.16	3.60
Task 4: Leadership for Family Engagement and Community Involvement	302	2.85	0.33	1.61	3.58

Table 5. Descriptive Statistics of Rubrics by Task for 2018–19 All Takers

		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	366	2.96	0.32	1.66	4.00
	Rubric 1b: Design an Integrated Plan for Strategies to Develop and Implement Improvement in the Priority Academic Area	366	2.92	0.37	1.00	4.00
	Rubric 1c: Assess and Analyze Feedback from Participants	366	2.90	0.37	1.00	4.00
Task 2: Instructional Leadership for a Professional Learning Culture	Rubric 2a: Plan to Facilitate Group Learning	323	2.96	0.25	2.00	3.75
	Rubric 2b: Enact a Professional Learning Culture to Support Team Learning	323	2.88	0.41	1.00	4.00
	Rubric 2c: Assess Team Learning to Improve Ongoing Group Learning	323	2.95	0.35	1.50	4.00
Task 3: Leadership in Observing, Assessing, and Supporting Individual Teacher Effectiveness	Rubric 3a: Plan	313	2.98	0.34	1.00	4.00
	Rubric 3b: Conduct the Observation	313	2.95	0.27	2.00	4.00
	Rubric 3c: Provide Feedback and Suggest Support	313	2.97	0.24	2.00	4.00
	Rubric 3d: Assess: Analyze and Identify Implications	313	2.87	0.36	2.00	4.00
Task 4: Leadership for Family Engagement and Community Involvement	Rubric 4a: Plan to Promote Family and Community Involvement	302	2.88	0.28	1.66	3.50
	Rubric 4b: Implement an Engagement or Involvement Strategy	302	2.88	0.45	1.00	4.00
	Rubric 4c: Analyze Feedback from Participants and Assess Leadership Skills	302	2.81	0.48	1.00	3.75

Table 6. Descriptive Statistics of Indicators by Task for 2018–19 PAL All Takers

			Indicator Score				
			N	Mean	S.D.	Min	Max
Task	Rubric	Indicator					
Task 1: Leadership through a Vision for High Student Achievement	Rubric 1a: Investigate and Prepare a Vision	Indicator 1a1: Data collection	366	2.97	0.34	1.50	4.00
		Indicator 1a2: Data analysis and priority definition	366	2.97	0.39	1.50	4.00
		Indicator 1a3: Evaluation of existing policies, practices, and programs	366	2.96	0.37	1.50	4.00
	Rubric 1b: Design an Integrated Plan for Strategies to Develop and Implement Improvement in the Priority Academic Area	Indicator 1b1: Vision plan and focus	366	2.93	0.42	1.00	4.00
		Indicator 1b2: Solicitation of input from teachers and other stakeholders	366	2.90	0.48	1.00	4.00
		Indicator 1b3: Plan details	366	2.94	0.43	1.00	4.00
	Rubric 1c: Assess and Analyze Feedback from Participants	Indicator 1c1: Plan feedback	366	2.98	0.46	1.00	4.00
		Indicator 1c2: Assessment of leadership skills and practices	366	2.82	0.40	1.00	4.00

			Indicator Score				
			N	Mean	S.D.	Min	Max
Task 2: Instructional Leadership for a Professional Learning Culture	Rubric 2a: Plan to Facilitate Group Learning	Indicator 2a1: Group identification and group formation	323	2.98	0.27	1.00	4.00
		Indicator 2a2: Group learning plan	323	2.93	0.34	2.00	4.00
	Rubric 2b: Enact a Professional Learning Culture to Support Team Learning	Indicator 2b1: Group process	323	2.97	0.37	1.00	4.00
		Indicator 2b2: Group learning and work	323	2.79	0.53	1.00	4.00
	Rubric 2c: Assess Team Learning to Improve Ongoing Group Learning	Indicator 2c1: Assessment of group process and group work	323	2.96	0.41	1.00	4.00
		Indicator 2c2: Assessment of leadership skills and practices	323	2.94	0.42	1.00	4.00
Task 3: Leadership in Observing, Assessing, and Supporting Individual Teacher Effectiveness	Rubric 3a: Plan	Indicator 3a1: Observation focus selection	313	3.01	0.41	1.00	4.00
		Indicator 3a2: Pre-observation meeting	313	2.94	0.37	1.00	4.00
	Rubric 3b: Conduct the Observation	Indicator 3b1: Use and application of teacher observation rubric	313	2.94	0.32	2.00	4.00
		Indicator 3b2: Description of observations	313	2.96	0.31	2.00	4.00
	Rubric 3c: Provide Feedback and Suggest Support	Indicator 3c1: Feedback content	313	2.93	0.27	2.00	4.00
		Indicator 3c2: Rapport and teacher engagement	313	3.08	0.34	2.00	4.00
		Indicator 3c3: Teacher development	313	2.91	0.36	1.00	4.00
	Rubric 3d: Assess: Analyze and Identify Implications	Indicator 3d1: Assessment of leadership skills and practices	313	2.87	0.36	2.00	4.00

			Indicator Score				
			N	Mean	S.D.	Min	Max
Task 4: Leadership for Family Engagement and Community Involvement	Rubric 4a: Plan to Promote Family and Community Involvement	Indicator 4a1: Investigation of the priority area	302	2.85	0.35	1.50	4.00
		Indicator 4a2: Investigation of work group engagement	302	2.94	0.35	1.50	4.00
		Indicator 4a3: Preparation of the plan, including strategies	302	2.85	0.38	1.00	4.00
	Rubric 4b: Implement an Engagement or Involvement Strategy	Indicator 4b1: Implementation of the strategy	302	2.88	0.45	1.00	4.00
	Rubric 4c: Analyze Feedback from Participants and Assess Leadership Skills	Indicator 4c1: Assessment and analysis of feedback on the family and community engagement plan and strategy	302	2.82	0.51	1.00	4.00
		Indicator 4c2: Assessment of leadership skills and practices	302	2.80	0.60	1.00	4.00

Massachusetts PAL 2018–2019 Technical Report

During the 2016–2019 Program Years, candidates were required to achieve a higher composite average score (2.75) to pass than they were during the initial 2015–2016 program year (2.5). For this reason, the percentage of PAL completers who passed (and how well) continues to be evaluated at critical score ranges.

In Program Year 2017–2018, 8.2 percent of completers had scores that did not meet or exceed the total average composite PAL score level (2.75). Similarly, evaluation results in Table 7 show that 8.8 percent of completers had scores that did not meet or exceeded the total average composite PAL score level for the 2018–2019 program year.

Table 7. Percentage Distribution of 2018–19 PAL Completers by Total Average PAL Score

Score Range	Number	Percent
Less than 2.50	11	3.6
2.50–2.74	16	5.2
2.75–2.99	187	60.9
3.00 and above	93	30.3
All Completers	307	100.0

The PAL Completers’ total PAL scores were compared by preparation pathway, gender, and race/ethnicity, and the results are shown in Table 8. The overall mean was 2.92 (comparable to the 2017–2018 program year) with a 0.18 standard deviation. Using t-tests for comparison of pairs by pathway and gender, the results were not statistically significant.

Table 8. Descriptive Statistics of Total Scores by Demographic Attributes of 2018–19 PAL Completers

	Total Score				
	N	Mean	S.D.	Min	Max
All Completers	307	2.92	0.18	2.09	3.38
Preparation Pathway					
Preparation Program	216	2.93	0.18	2.09	3.38
Alternative Pathway— Internship	78	2.88	0.17	2.23	3.24
Alternative Pathway— Panel Review	12	2.79	0.19	2.42	3.00
Out of State	--	--	--	--	--
Gender					
Male	88	2.94	0.18	2.41	3.33
Female	218	2.91	0.18	2.09	3.38

Massachusetts PAL 2018–2019 Technical Report

	Total Score				
	N	Mean	S.D.	Min	Max
Not Reported	--	--	--	--	--
Race/Ethnicity					
Native American	0	--	--	--	--
Asian	--	--	--	--	--
Black	13	2.86	0.10	2.74	3.05
White	262	2.92	0.18	2.09	3.38
Hispanic	--	--	--	--	--
Native American/Latino	0	--	--	--	--
Asian/Latino	--	--	--	--	--
Black/Latino	--	--	--	--	--
Pacific	--	--	--	--	--
Pacific/Latino	0	--	--	--	--
White/Latino	--	--	--	--	--
Multiracial	--	--	--	--	--
Multiracial/Latino	--	--	--	--	--
Not reported	--	--	--	--	--

Next, we examined patterns of demographic attributes and candidate performance for all candidates. All scores showed no statistically significant difference. As shown in Table 9, in Program Year 2018–2019 there was a pattern in the differences between tasks by gender but not Task 3 in preparation pathway. Male and female PAL candidate task scores were the same for Task 3 (2.94); however, male PAL candidate scores were higher for Tasks 1, 2, and 4.

Table 9. Best Attempt Descriptive Statistics by Demographic Characteristic for All Takers Submitting Tasks in 2018–2019

	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
Candidates	366	2.93	0.31	1.33	3.66	323	2.92	0.29	1.66	3.83	313	2.94	0.21	2.16	3.60	302	2.85	0.33	1.61	3.58
Preparation Pathway																				
Preparation Program	265	2.95	0.29	1.33	3.66	229	2.95	0.27	1.75	3.83	220	2.94	0.21	2.16	3.60	215	2.87	0.34	1.61	3.58
Alternative Pathway— Internship	86	2.86	0.32	1.77	3.61	81	2.86	0.32	1.66	3.41	80	2.96	0.19	2.25	3.45	75	2.82	0.34	1.77	3.52
Alternative Pathway— Panel Review	12	2.84	0.39	1.83	3.41	11	2.75	0.36	2.16	3.08	12	2.82	0.23	2.43	3.18	11	2.76	0.29	2.05	3.08
Out of State	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Gender																				
Male	112	2.95	0.28	2.00	3.66	100	2.95	0.29	2.16	3.83	96	2.94	0.21	2.16	3.60	89	2.88	0.35	1.66	3.52
Female	252	2.91	0.32	1.33	3.50	222	2.91	0.29	1.66	3.58	216	2.94	0.21	2.25	3.50	212	2.84	0.33	1.61	3.58
Not Reported	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Race/Ethnicity																				
Native American	0	--	--	--	--	0	--	--	--	--	0	--	--	--	--	0	--	--	--	--
Asian	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Black	13	2.84	0.23	2.30	3.22	15	2.86	0.39	2.16	3.41	19	2.92	0.29	2.25	3.47	16	2.84	0.28	2.38	3.58
White	314	2.93	0.31	1.33	3.66	276	2.92	0.28	1.66	3.83	260	2.94	0.21	2.16	3.60	251	2.86	0.34	1.61	3.58
Hispanic	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Native American/Latino	0	--	--	--	--	0	--	--	--	--	0	--	--	--	--	0	--	--	--	--
Asian/Latino	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Massachusetts PAL 2018–2019 Technical Report

	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
Black/Latino	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Pacific	0	--	--	--	--	0	--	--	--	--	0	--	--	--	--	--	--	--	--	--
Pacific/Latino	0	--	--	--	--	0	--	--	--	--	0	--	--	--	--	0	--	--	--	--
White/Latino	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Multiracial	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Multiracial/Latino	--	--	--	--	--	0	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Not reported	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Task scores for the 2018–2019 program year were correlated to evaluate the degree of association. As was the case for the 2017–2018 program year, the four factors have a positive correlation for Task 1, 2, 3, and 4. Although Task 2 and Task 4 held the highest task-based correlation for Program Year 2017–2018 (0.52039, N=194), this program year Task 2 and Task 3 hold the highest task-based correlation (0.38666, N=253).

Table 10. Factor Correlation for 2017–18 PAL Tasks

Pearson Correlation Coefficients (N)				
	Task 1	Task 2	Task 3	Task 4
Task 1	1.00000 (366)	0.27251 <.0001 (263)	0.34886 <.0001 (231)	0.35124 <.0001 (228)
Task 2	0.27251 <.0001 (263)	1.00000 (323)	0.38666 <.0001 (253)	0.33879 <.0001 (258)
Task 3	0.34886 <.0001 (231)	0.38666 <.0001 (253)	1.00000 (313)	0.35702 <.0001 (254)
Task 4	0.35124 <.0001 (228)	0.33879 <.0001 (258)	0.35702 <.0001 (254)	1.00000 (302)

Note: All correlation coefficients are significant at the $p < .0001$ level. N are indicated in parentheses.

The PAL Completer score results are positive for Program Year 2018–19, with an overall pass rate of 91%. This performance is comparable to prior program years’ performance. Pass rates differ by preparation pathway for Tasks 1, 2, and 4, with Task 4 having the widest range (91% to 99%). Task 1, 2, 3, and Task 4 pass rates for all takers by gender are the same or differ by only 2 to 3%.

Compared with the number of Completers for Program Year 2017–2018 (N=221), the increased number of Completers for Program Year 2018–2019 (N=307) strengthened the statistical comparison of candidate performance and evaluation of the scores through factor analysis and reliability analyses. Correlations of candidate scores between tasks show that these tasks continue to be independent measures with modest degrees of association, with the strongest being between Tasks 2 and 3 and the weakest between Tasks 1 and 2.

Cronbach’s alpha is a measure of internal consistency of raw test scores, an important characteristic of test scores that indicates the extent to which the items of the assessment measure the intended common construct (Cronbach, 1951). Cronbach’s alpha estimates range from zero to one, and higher values reflect higher levels of consistency of a person’s scores across the items (rubrics). Task-level Alpha Estimates are presented in Table 11.

Table 11. Alpha by Task

Task Level Alpha Estimates	
Task	Alpha
1	0.65095
2	0.62715
3	0.60376
4	0.73209

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) were calculated for each indicator.

Table 12 presents the results of the rater agreement and kappa n calculated using percent exact only. 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 that kappa n ranged from 0.45 (indicator 2b.2) to 0.77 (indicator 2a.1). Exact agreement rates, indicating the percentage of cases where scorers scoring the same portfolio assigned the same score, are at or above 59% on all indicators.

In addition to examining rater agreement at exact and adjacent, agreement was also reviewed by indicator as a grouping to see the percent of agreement where both scorer 1 and 2 were either at or below 2 or at or above 3. This directionality helps to identify indicators that may not otherwise appear to clearly discern between critical score points, and indicators where there is more of a disparity between raters. Agreement rates, as shown in Table 12, indicate that independent scorers scoring the same submission tend to assign the same directionality (both scores are 2 or below or both scores are 3 or above) more often than they agree exactly. Rubric 2b.2 has the lowest combination of agreement between both exact agreement and directional agreement.

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

Indicator	N of Items Scored	Rater Agreement		Inter-Rater Reliability	Kappa N with Exact Agreement	Percent (%)
		Percent Exact	Percent Exact and Adjacent			Both Scorers ≤2 or ≥3
1a.1	370	0.81	0.99	0.54	0.74	0.92
1a.2	370	0.76	0.99	0.55	0.68	0.92

Massachusetts PAL 2018–2019 Technical Report

	N of Items Scored	Rater Agreement		Inter-Rater Reliability	Kappa N with Exact Agreement	Percent (%)
		Percent Exact	Percent Exact and Adjacent			Both Scorers ≤ 2 or ≥ 3
1a.3	370	0.74	0.98	0.41	0.65	0.87
1b.1	370	0.68	1.00	0.54	0.57	0.82
1b.2	370	0.60	0.96	0.35	0.47	0.77
1b.3	370	0.64	0.99	0.47	0.52	0.81
1c.1	370	0.62	0.99	0.49	0.49	0.83
1c.2	370	0.73	0.99	0.53	0.64	0.79
2a.1	334	0.83	1.00	0.41	0.77	0.90
2a.2	334	0.72	0.99	0.39	0.62	0.81
2b.1	334	0.75	1.00	0.57	0.67	0.89
2b.2	334	0.59	0.98	0.59	0.45	0.72
2c.1	334	0.67	0.97	0.45	0.56	0.82
2c.2	334	0.65	1.00	0.50	0.53	0.80
3a.1	314	0.72	0.99	0.50	0.62	0.88
3a.2	314	0.77	1.00	0.57	0.69	0.86
3b.1	314	0.76	0.99	0.36	0.69	0.86
3b.2	314	0.79	1.00	0.44	0.72	0.87
3c.1	314	0.79	1.00	0.28	0.72	0.85
3c.2	314	0.68	0.99	0.20	0.58	0.92
3c.3	314	0.74	1.00	0.49	0.65	0.81
3d.1	314	0.75	1.00	0.56	0.67	0.80
4a.1	327	0.76	0.98	0.52	0.68	0.80
4a.2	327	0.75	0.99	0.57	0.66	0.84
4a.3	327	0.76	0.99	0.63	0.68	0.82
4b.1	327	0.69	0.98	0.62	0.58	0.82
4c.1	327	0.66	0.99	0.69	0.55	0.80
4c.2	327	0.60	0.98	0.67	0.47	0.76

To further explore the distribution of scores assigned to each indicator, four x four contingency tables were created for each indicator and are found in Appendix A. These tables illustrate, for each indicator, the distribution of score points assigned by both scorers across the 1 to 4 scale. These tables provide detailed information on the percent agreement at each combination of score points.

Additionally, rater agreement was examined at the critical threshold of passing/failing (Task Score of 2.1), in order to explore decision consistency by task between scorer 1 and scorer 2. Table 13 indicates the percent agreement between Scorer 1 and Scorer 2’s total indicator scores in relation to the total task score of 2.1—outlining the consistency of a final task outcome of pass or fail. Consistency ranges from 93% (Task 4) to 98% (Task 3). These agreements are before any adjudication by a scoring supervisor, and the scoring model does indicate that 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.

Table 13. Rater Agreement and Decision Consistency by Task

Task 1 Decision Consistency Between Scorers at Task Score 2.1 (%)		
	Scorer 1 Pass	Scorer 1 Fail
Scorer 2 Pass	96	1
Scorer 2 Fail	1	2

Task 2 Decision Consistency Between Scorers at Task Score 2.1 (%)		
	Scorer 1 Pass	Scorer 1 Fail
Scorer 2 Pass	95	3
Scorer 2 Fail	2	0

Task 3 Decision Consistency Between Scorers at Task Score 2.1 (%)		
	Scorer 1 Pass	Scorer 1 Fail
Scorer 2 Pass	100	0
Scorer 2 Fail	0	0

Task 4 Decision Consistency Between Scorers at Task Score 2.1 (%)		
	Scorer 1 Pass	Scorer 1 Fail
Scorer 2 Pass	89	3
Scorer 2 Fail	6	2

References

- Brennan, R. L., & Prediger, D. J. (1981). Coefficient Kappa: Some uses, misuses, and alternatives. *Educational and Psychological Measurement, 41*(3), 687–699.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297–334.
- 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.

Appendix A

Appendix A: Score Distributions (Percent Agreement) between Scorer 1 and Scorer 2, by Indicator

TASK 1

Joint Score Distribution for Task 1 Indicator 1a1 (%)					
Scorer 1	4	0	0	4.59	0.81
	3	0	5.14	75.41	5.68
	2	0.54	4.32	2.7	0.54
	1	0	0.27	0	0
		1	2	3	4
		Scorer 2			

Joint Score Distribution for Task 1 Indicator 1a2 (%)					
Scorer 1	4	0	0.27	8.11	1.35
	3	0	2.97	70	6.22
	2	0.54	4.86	4.32	0.54
	1	0	0.81	0	0
		1	2	3	4
		Scorer 2			

Joint Score Distribution for Task 1 Indicator 1a3 (%)					
Scorer 1	4	0	0	5.95	1.35
	3	1.08	4.59	69.46	6.22
	2	0	3.24	7.03	0.27
	1	0	0.54	0.27	0
		1	2	3	4
		Scorer 2			

Joint Score Distribution for Task 1 Indicator 1b1 (%)					
Scorer 1	4	0	0.27	6.76	2.43
	3	0	8.65	59.73	5.95
	2	0.54	5.95	9.19	0
	1	0	0.54	0	0
		1	2	3	4
		Scorer 2			

Joint Score Distribution for Task 1 Indicator 1b2 (%)					
Scorer 1	4	0	1.89	7.3	1.35
	3	0.81	8.65	53.24	8.65
	2	0	5.41	10.54	0.54
	1	0.27	1.08	0.27	0
		1	2	3	4
Scorer 2					

Joint Score Distribution for Task 1 Indicator 1b3 (%)					
Scorer 1	4	0	0.27	8.11	1.62
	3	0	11.89	55.68	8.11
	2	0	6.22	6.76	0.54
	1	0.27	0.54	0	0
		1	2	3	4
Scorer 2					

Joint Score Distribution for Task 1 Indicator 1c1 (%)					
Scorer 1	4	0	0.81	9.19	2.7
	3	0	7.57	55.14	10.54
	2	0.81	3.78	8.11	0
	1	0.27	0.81	0.27	0
		1	2	3	4
Scorer 2					

Joint Score Distribution for Task 1 Indicator 1c2 (%)					
Scorer 1	4	0	0	2.97	0.27
	3	0	9.73	64.05	1.89
	2	0.54	8.65	10.27	0.54
	1	0.27	0.54	0.27	0
		1	2	3	4
Scorer 2					

TASK 2

Joint Score Distribution for Task 2 Indicator 2a1 (%)					
Scorer 1	4	0	0	4.79	0.3
	3	0	4.79	79.94	2.69
	2	0	2.1	4.79	0.3
	1	0.3	0	0	0
		1	2	3	4
Scorer 2					

Joint Score Distribution for Task 2 Indicator 2a2 (%)					
Scorer 1	4	0	0	4.49	0.9
	3	0.3	9.28	65.87	4.49
	2	0	5.09	9.28	0.3
	1	0	0	0	0
		1	2	3	4
Scorer 2					

Joint Score Distribution for Task 2 Indicator 2b1 (%)					
Scorer 1	4	0	0	6.89	0.6
	3	0.3	4.79	68.86	5.99
	2	0.3	5.39	5.99	0
	1	0.3	0.6	0	0
		1	2	3	4
Scorer 2					

Joint Score Distribution for Task 2 Indicator 2b2 (%)					
Scorer 1	4	0	0.6	4.19	3.59
	3	0	13.17	40.42	5.99
	2	1.5	14.37	12.87	1.2
	1	0.6	0.9	0.6	0
		1	2	3	4
Scorer 2					

Joint Score Distribution for Task 2 Indicator 2c1 (%)					
Scorer 1	4	0	0.6	4.49	1.8
	3	0.3	7.19	58.68	9.88
	2	0.3	6.59	7.49	1.2
	1	0.3	0.3	0.9	0
		1	2	3	4
Scorer 2					

Massachusetts PAL 2018–2019 Technical Report

Joint Score Distribution for Task 2 Indicator 2c2 (%)					
Scorer 1	4	0	0	5.69	1.8
	3	0	9.58	56.89	9.28
	2	0.3	5.69	10.48	0
	1	0.3	0	0	0
		1	2	3	4
		Scorer 2			

TASK 3

Joint Score Distribution for Task 3 Indicator 3a1 (%)					
Scorer 1	4	0	0.64	7.01	3.18
	3	0	6.05	64.33	9.55
	2	0	3.82	5.1	0
	1	0.32	0	0	0
		1	2	3	4
Scorer 2					

Joint Score Distribution for Task 3 Indicator 3a2 (%)					
Scorer 1	4	0	0	4.78	1.59
	3	0	5.1	71.02	4.46
	2	0	3.82	8.6	0
	1	0.64	0	0	0
		1	2	3	4
Scorer 2					

Joint Score Distribution for Task 3 Indicator 3b1 (%)					
Scorer 1	4	0	0.32	3.82	0.64
	3	0	5.1	72.61	5.1
	2	0	3.18	8.6	0.32
	1	0	0.32	0	0
		1	2	3	4
Scorer 2					

Joint Score Distribution for Task 3 Indicator 3b2 (%)					
Scorer 1	4	0	0	2.55	1.59
	3	0	5.41	74.52	5.1
	2	0	2.87	7.64	0
	1	0	0	0.32	0
		1	2	3	4
Scorer 2					

Joint Score Distribution for Task 3 Indicator 3c1 (%)					
Scorer 1	4	0	0	1.27	0.32
	3	0	4.78	76.11	4.46
	2	0	2.55	10.19	0.32
	1	0	0	0	0
		1	2	3	4
Scorer 2					

Joint Score Distribution for Task 3 Indicator 3c2 (%)					
Scorer 1	4	0	0	9.24	1.91
	3	0	1.27	64.01	14.33
	2	0	2.23	5.73	1.27
	1	0	0	0	0
		1	2	3	4
Scorer 2					

Joint Score Distribution for Task 3 Indicator 3c3 (%)					
Scorer 1	4	0	0.32	3.18	1.59
	3	0	7.96	67.2	4.14
	2	0	4.78	10.51	0
	1	0.32	0	0	0
		1	2	3	4
Scorer 2					

Joint Score Distribution for Task 3 Indicator 3d1 (%)					
Scorer 1	4	0	0	3.82	1.27
	3	0	7.64	65.29	1.27
	2	0	8.6	12.1	0
	1	0	0	0	0
		1	2	3	4
Scorer 2					

TASK 4

Joint Score Distribution for Task 4 Indicator 4a1 (%)					
Scorer 1	4	0	0	1.53	0.31
	3	0.61	10.4	67.58	0.61
	2	0.92	7.95	8.26	0
	1	0	0.92	0.92	0
		1	2	3	4
Scorer 2					

Joint Score Distribution for Task 4 Indicator 4a2 (%)					
Scorer 1	4	0	0	3.98	0.92
	3	0.31	8.26	67.89	3.67
	2	0.92	5.81	7.03	0
	1	0	0.92	0.31	0
		1	2	3	4
Scorer 2					

Joint Score Distribution for Task 4 Indicator 4a3 (%)					
Scorer 1	4	0	0.31	2.14	0
	3	0.31	7.95	64.53	2.45
	2	0.92	11.31	9.17	0
	1	0.31	0.61	0	0
		1	2	3	4
Scorer 2					

Joint Score Distribution for Task 4 Indicator 4b1 (%)					
Scorer 1	4	0	0.61	4.28	1.22
	3	0.61	6.42	53.82	6.42
	2	0.61	13.46	9.79	0
	1	0.31	1.53	0.92	0
		1	2	3	4
Scorer 2					

Joint Score Distribution for Task 4 Indicator 4c1 (%)					
Scorer 1	4	0	0.31	5.5	1.22
	3	0	10.09	48.01	5.5
	2	1.53	15.29	9.17	0.61
	1	1.53	1.22	0	0
		1	2	3	4
Scorer 2					

Joint Score Distribution for Task 4 Indicator 4c2 (%)					
Scorer 1	4	0	0.92	7.34	3.67
	3	0	10.7	37.92	5.2
	2	1.22	18.04	11.31	0.92
	1	0.61	2.14	0	0
		1	2	3	4
		Scorer 2			