MASSACHUSETTS

PERFORMANCE ASSESSMENT *for* LEADERS

Annual Technical Report

2017–2018 Program Year

Massachusetts Department of
ELEMENTARY & SECONDARY
EDUCATION

Contents

Introduction

This report presents the results for the third program year (2017–2018) 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 PAL 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, and all three require a candidate to pass the PAL assessment. These are described 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 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' 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).

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 also applied for the 2017–2018 operational program year.

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., "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 for the 2017–2018 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 2017–2018 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 2017–2018 program year. 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 2017–2018 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 2018. 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 2017–2018

Following presents information on performance during the 2017–2018 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 2017–2018 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 July 28, 2017 and June 14, 2018). 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 14, 2018. Reports based on Completers' data, therefore, include best attempts only.

All Takers: Candidates who have submitted at least one scorable task between July 28, 2017 and June 14, 2018.

A summary of candidate numbers is as follows:

1. The N of PAL Completers is 147; i.e., 147 candidates completed their final task in the 2017–2018 program year, thereby obtaining scores on all four tasks within the PAL Assessment.

2. The N of All Takers (2017–2018) attempting at least one task is 379, as follows:

Table 1. Number of Tasks Completed by PAL Candidates

N of Tasks Completed	N of Candidates
4	147
3	59
2	61
1	112
Total	379

During Program Year 2017–18, 379 individual candidates submitted to Pearson at least one scorable PAL portfolio task toward the assessment for licensure (as shown in Table 1). This included 147 candidates who completed all four tasks, 59 candidates who completed three tasks, 61 candidates who completed two tasks, and 112 candidates who completed one task. In contrast, 323 individual candidates submitted at least one scorable PAL portfolio task and 151 candidates completed all four tasks in Program Year 2016–17.

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 the PAL assessment by submitting their final task and passing the PAL in the 2017–2018 program year. This includes both individuals who began submitting tasks in the 2016–2017 program year, but only passes upon submitting their final task in the 2017–2018 program year. The first column also includes candidates who submitted all four tasks and passed the PAL assessment during the 2017–2018 program year only. The overall pass rate for the PAL assessment during the 2017–2018 program year was 92% (N=221).

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 higher rate than candidates through traditional preparation programs, although the N difference reflects about four times more candidates in the traditional preparation program route. Twice as many females completed the assessment as males, although both groups had similar pass rates. However, there was a difference between pass rates for candidates submitting Task 2 in the 2017–2018 program year, with a 99% pass rate for females in contrast to the 93% pass rate for males.

Table 2. Best Attempt Assessment Pass Rates by Demographic Characteristic for Candidates Submitting Tasks in 2017–2018

	Assessment (Completed Final Task During 2017–2018)		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	221	92%	293	97%	225	97%	228	99%	253	96%
Preparation Pathway										
Preparation Program	170	91%	221	97%	166	98%	170	99%	192	96%
Alternative Pathway-Internship	44	95%	60	97%	49	98%	50	100%	50	98%
Alternative Pathway-Panel Review	- -	- -	10	80%	- -	- -	- -	- -	- -	- -
Out of State	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
Gender										
Male	72	92%	87	97%	71	93%	69	99%	80	95%
Female	147	92%	204	97%	152	99%	157	99%	171	96%
Not Reported	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
Race/Ethnicity										
Native American	0	0%	0	0%	0	0%	0	0%	0	0%
Asian	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
Black	10	90%	16	88%	14	86%	- -	- -	15	87%
White	187	91%	251	97%	186	99%	196	99%	216	97%

	Assessment (Completed Final Task During 2017–2018)		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	0	0%	0	0%	0	0%	0	0%	0	0%
Black/Latino	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
Pacific	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
Pacific/Latino	0	0%	0	0%	0	0%	0	0%	0	0%
White/Latino	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
Multiracial	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
Not reported	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -

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 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 2017–2018 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 2017–2018

	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	293	3	22	74	1	225	3	14	82	1
Preparation Pathway										
Preparation Program	221	3	20	76	1	166	2	12	84	2
Alternative Pathway-Internship	60	3	30	67	0	49	2	24	73	0
Alternative Pathway-Panel Review	10	20	10	70	0	- -	- -	- -	- -	- -
Out of State	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
Gender										
Male	87	3	21	76	0	71	7	14	75	4
Female	204	3	23	73	1	152	1	14	85	0
Not Reported	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
Race/Ethnicity										
Native American	0	- -	- -	- -	- -	0	- -	- -	- -	- -
Asian	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
Black	16	13	38	50	0	14	14	21	64	0
White	251	3	21	76	0	186	1	14	83	2
Hispanic	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
Native American/Latino	0	- -	- -	- -	- -	0	- -	- -	- -	- -
Asian/Latino	0	- -	- -	- -	- -	0	- -	- -	- -	- -

	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	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
Pacific/Latino	0	- -	- -	- -	- -	0	- -	- -	- -	- -
White/Latino	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
Multiracial	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
Multiracial/Latino	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
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	228	1	15	83	1	253	4	26	70	0
Preparation Pathway										
Preparation Program	170	1	16	81	1	192	4	23	72	0
Alternative Pathway-Internship	50	0	8	92	0	50	2	36	62	0
Alternative Pathway-Panel Review	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
Out of State	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
Gender										
Male	69	1	14	84	0	80	5	16	79	0

	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	157	1	15	83	1	171	4	31	65	0
Race/Ethnicity										
NatAmer	0	- -	- -	- -	- -	0	- -	- -	- -	- -
Asian	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
Black	- -	- -	- -	- -	- -	15	13	33	53	0
White	196	1	14	84	1	216	3	28	69	0
Hispanic	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
NatAmer/Latino	0	- -	- -	- -	- -	0	- -	- -	- -	- -
Asian/Latino	0	- -	- -	- -	- -	0	- -	- -	- -	- -
Black/Latino	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
Pacific	- -	- -	- -	- -	- -	0	- -	- -	- -	- -
Pacific/Latino	0	- -	- -	- -	- -	0	- -	- -	- -	- -
White/Latino	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
Multiracial	- -	- -	- -	- -	- -	0	- -	- -	- -	- -
Multiracial/Latino	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
Not reported	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -

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. As was the case in the 2016–2017 program year, mean task scores are higher for Tasks 2 and 3, with Task 4 having a lower mean performance and slightly larger standard deviation.

Table 5 shows that average rubric scores range from 2.70 up to 3.00. The highest performing rubrics were 2a (Plan to Facilitate Group Learning) and 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 2.64 to 3.04. The highest performing indicator was 3c2 (Rapport and teacher engagement) and the lowest performing indicator was 4c2 (Assessment of leadership skills and practices). These were also the highest and lowest performing indicators in the 2016–2017 program year. Note that one submission by a candidate was unscorable for Indicator 2c2 (Assessment of leadership skills and practices).

Table 4. Descriptive Statistics of Tasks for 2017–18 All Takers

Task	Task Score				
	N	Mean	S.D.	Min	Max
Task 1: Leadership through a Vision for High Student Achievement	293	2.85	0.32	1.38	3.50
Task 2: Instructional Leadership for a Professional Learning Culture	225	2.94	0.32	1.25	3.91
Task 3: Leadership in Observing, Assessing, and Supporting Individual Teacher Effectiveness	228	2.94	0.25	1.87	3.70
Task 4: Leadership for Family Engagement and Community Involvement	253	2.80	0.34	1.00	3.41

Table 5. Descriptive Statistics of Rubrics by Task for 2017–18 All Takers

Task	Rubric	N	Mean	S.D.	Min	Max
				Rubric Score		
Task 1: Leadership through a Vision for High Student Achievement	**Rubric 1a: Investigate and Prepare a Vision**	293	2.90	0.33	1.33	4.00
	Rubric 1b: Design an Integrated Plan for Strategies to Develop and Implement Improvement in the Priority Academic Area	293	2.86	0.39	1.00	3.83
	Rubric 1c: Assess and Analyze Feedback from Participants	293	2.80	0.37	1.25	3.75
Task 2: Instructional Leadership for a Professional Learning Culture	**Rubric 2a: Plan to Facilitate Group Learning**	225	3.00	0.29	1.25	4.00
	Rubric 2b: Enact a Professional Learning Culture to Support Team Learning	225	2.89	0.42	1.00	4.00
	Rubric 2c: Assess Team Learning to Improve Ongoing Group Learning	225	2.93	0.39	1.50	4.00
Task 3: Leadership in Observing, Assessing, and Supporting Individual Teacher Effectiveness	**Rubric 3a: Plan**	228	3.00	0.39	1.50	4.00
	Rubric 3b: Conduct the Observation	228	2.98	0.25	2.00	4.00
	Rubric 3c: Provide Feedback and Suggest Support	228	2.98	0.29	2.00	3.83
	Rubric 3d: Assess: Analyze and Identify Implications	228	2.83	0.41	2.00	4.00
Task 4: Leadership for Family Engagement and Community Involvement	**Rubric 4a: Plan to Promote Family and Community Involvement**	253	2.85	0.32	1.00	3.66
	Rubric 4b: Implement an Engagement or Involvement Strategy	253	2.87	0.44	1.00	4.00
	Rubric 4c: Analyze Feedback from Participants and Assess Leadership Skills	253	2.70	0.44	1.00	3.75

Table 6. Descriptive Statistics of Indicators by Task for 2017–18 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**	293	2.91	0.34	1.50	4.00
		Indicator 1a2: Data analysis and priority definition	293	2.89	0.37	1.00	4.00
		Indicator 1a3: Evaluation of existing policies, practices, and programs	293	2.90	0.40	1.00	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**	293	2.88	0.47	1.00	4.00
		Indicator 1b2: Solicitation of input from teachers and other stakeholders	293	2.84	0.45	1.00	3.50
		Indicator 1b3: Plan details	293	2.86	0.45	1.00	4.00
	Rubric 1c: Assess and Analyze Feedback from Participants	**Indicator 1c1: Plan feedback**	293	2.89	0.43	1.00	4.00
		Indicator 1c2: Assessment of leadership skills and practices	293	2.72	0.43	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	225	3.02	0.30	1.50	4.00
		Indicator 2a2: Group learning plan	225	2.98	0.35	1.00	4.00
	Rubric 2b: Enact a Professional Learning Culture to Support Team Learning	Indicator 2b1: Group process	225	2.97	0.39	1.00	4.00
		Indicator 2b2: Group learning and work	225	2.82	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	225	2.93	0.47	1.00	4.00
		Indicator 2c2: Assessment of leadership skills and practices	224	2.94	0.41	2.00	4.00
Task 3: Leadership in Observing, Assessing, and Supporting Individual Teacher Effectiveness	Rubric 3a: Plan	Indicator 3a1: Observation focus selection	238	2.99	0.51	1.00	4.00
		Indicator 3a2: Pre-observation meeting	238	2.97	0.43	1.00	4.00
	Rubric 3b: Conduct the Observation	Indicator 3b1: Use and application of teacher observation rubric	238	3.00	0.34	2.00	4.00
		Indicator 3b2: Description of observations	238	2.96	0.28	2.00	4.00
	Rubric 3c: Provide Feedback and Suggest Support	Indicator 3c1: Feedback content	238	2.99	0.34	2.00	4.00
		Indicator 3c2: Rapport and teacher engagement	238	3.04	0.41	2.00	4.00
		Indicator 3c3: Teacher development	238	2.88	0.40	2.00	4.00
	Rubric 3d: Assess: Analyze and Identify Implications	Indicator 3d1: Assessment of leadership skills and practices	238	2.80	0.42	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	269	2.81	0.40	1.00	3.50
		Indicator 4a2: Investigation of work group engagement	269	2.88	0.41	1.00	4.00
		Indicator 4a3: Preparation of the plan, including strategies	269	2.78	0.44	1.00	3.50
	Rubric 4b: Implement an Engagement or Involvement Strategy	Indicator 4b1: Implementation of the strategy	269	2.82	0.48	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	269	2.70	0.53	1.00	4.00
		Indicator 4c2: Assessment of leadership skills and practices	269	2.64	0.49	1.00	4.00

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. Eight percent had scores that did not meet or exceed the total average composite PAL score level (2.75) in Program Year 2016–2017. Similarly, evaluation results in Table 7 show that only 8.2 percent had scores that did not meet or exceeded the total average composite PAL score level for the 2017–2018 program year.

Table 7. Percentage Distribution of 2017–18 PAL Completers by Total Average PAL Score

Score Range	Number	Percent
Less than 2.50	5	2.3
2.50–2.74	13	5.9
2.75–2.99	131	59.3
3.00 and above	72	32.6
All Completers	221	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 2.91 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 2017–18 PAL Completers

	Total Score				
	N	**Mean**	**S.D.**	**Min**	**Max**
All Completers	221	2.91	0.18	1.84	3.35
Preparation Pathway					
Preparation Program	170	2.91	0.19	1.84	3.35
Alternative Pathway-Internship	44	2.91	0.13	2.72	3.31
Alternative Pathway-Panel Review	- -	- -	- -	- -	- -
Out of State	- -	- -	- -	- -	- -
Gender					
Male	72	2.91	0.22	1.84	3.35
Female	147	2.92	0.16	2.19	3.27
Race/Ethnicity					
Native American	0	- -	- -	- -	- -

	Total Score				
	N	**Mean**	**S.D.**	**Min**	**Max**
Asian	- -	- -	- -	- -	- -
Black	10	2.89	0.14	2.58	3.06
White	187	2.91	0.19	1.84	3.35
Hispanic	- -	- -	- -	- -	- -
Native American/Latino	0	- -	- -	- -	- -
Asian/Latino	0	- -	- -	- -	- -
Black/Latino	- -	- -	- -	- -	- -
Pacific	0	- -	- -	- -	- -
Pacific/Latino	0	- -	- -	- -	- -
White/Latino	- -	- -	- -	- -	- -
Multiracial	- -	- -	- -	- -	- -
Multiracial/Latino	- -	- -	- -	- -	- -
Not reported	- -	- -	- -	- -	- -

Next, we examined patterns of demographic attributes and candidate performance. For one individual in a race/ethnicity group with an N<10, there were statistically significant differences for scores on Tasks 2 and 4. All other scores showed no statistically significant difference. As shown in Table 9, in Program Year 2017–2018 there was a pattern in the differences between tasks by gender but not preparation pathway. Male and female PAL candidate task scores were the same for Task 1 (2.85) and approximately the same for Task 4 (2.81 and 2.80, respectively). However, female PAL candidate scores were higher for both Task 2 and Task 3.

Table 9. Best Attempt Descriptive Statistics by Demographic Characteristic for All Takers Submitting Tasks in 2017–2018

	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	293	2.85	0.31	1.50	3.50	225	2.94	0.32	1.25	3.91	228	2.94	0.25	1.87	3.70	253	2.80	0.34	1.00	3.41
Preparation Pathway																				
Preparation Program	221	2.87	0.30	1.50	3.50	166	2.96	0.30	1.41	3.91	170	2.94	0.25	1.87	3.70	192	2.81	0.35	1.00	3.38
Alternative Pathway-Internship	60	2.79	0.31	1.83	3.36	49	2.88	0.32	1.50	3.41	50	2.97	0.20	2.18	3.39	50	2.81	0.31	1.50	3.41
Alternative Pathway-Panel Review	10	2.71	0.49	1.88	3.22	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
Out of State	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
Gender																				
Male	87	2.85	0.30	1.50	3.47	71	2.90	0.45	1.25	3.91	69	2.91	0.24	1.87	3.39	80	2.81	0.36	1.00	3.33
Female	204	2.85	0.32	1.55	3.50	152	2.96	0.25	1.50	3.41	157	2.96	0.25	1.87	3.70	171	2.80	0.34	1.27	3.41
Not Reported	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
Race/Ethnicity																				
Native American	0	- -	- -	- -	- -	0	- -	- -	- -	- -	0	- -	- -	- -	- -	0	- -	- -	- -	- -
Asian	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
Black	16	2.63	0.37	1.75	3.00	14	2.72	0.63	1.25	3.41	- -	- -	- -	- -	- -	15	2.63	0.49	1.27	3.00
White	251	2.86	0.30	1.50	3.50	186	2.97	0.27	2.00	3.91	196	2.95	0.25	1.87	3.70	216	2.80	0.33	1.00	3.41
Hispanic	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
Native American/Latino	0	- -	- -	- -	- -	0	- -	- -	- -	- -	0	- -	- -	- -	- -	0	- -	- -	- -	- -
Asian/Latino	0	- -	- -	- -	- -	0	- -	- -	- -	- -	0	- -	- -	- -	- -	0	- -	- -	- -	- -

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

Task scores for the 2017–2018 program year were correlated to evaluate the degree of association. As was the case for the 2016–2017 program year, the four factors have a positive correlation for Task 1, 2, 3, and 4. Although Task 1 and Task 4 held the highest positive task-based correlation for Program Year 2016–2017 (0.42166, N=151), Task 2 and Task 4 hold the highest task-based correlation for Program Year 2017–2018 (0.52039, N=194).

Table 10. Factor Correlation for 2017–18 PAL Tasks

	Pearson Correlation Coefficients			
	Task 1	**Task 2**	**Task 3**	**Task 4**
Task 1	1.00000	0.39081	0.41839	0.35722
	(293)	(200)	(183)	(192)
Task 2	0.39081	1.00000	0.32319	0.52039
	(200)	(225)	(172)	(194)
Task 3	0.41839	0.32319	1.00000	0.34400
	(183)	(172)	(228)	(179)
Task 4	0.35722	0.52039	0.34400	1.00000
	(192)	(194)	(179)	(253)

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 2017–18, with an overall pass rate of 92%. This shows slight improvement over the Program Year 2016–2017 pass rate of 91%. Pass rates differ by preparation pathway for Task 4 all takers, but pass rates are the same for Task 1 and Task 2; Task 3 pass rates differ by only 1%. Task 1, Task 3, and Task 4 pass rates for all takers by gender are the same or differ by only 1%, while the range of pass rates for Task 2 is wider (93%–99%).

Compared with the number of completers for Program Year 2016–2017 (N=151), the increased number of completers for Program Year 2017–2018 (N=221) 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 4 and the weakest between Tasks 2 and 3.

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	0.854
2	0.845
3	0.682
4	0.826

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 0.44 (indicator 1b.2) to 0.80 (indicator 3b.2).

Exact agreement rates, indicating the percentage of cases where scorers scoring the same portfolio assigned the same score, are above 50 percent on all indicators.

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

		Rater Agreement			
	N of Items Scored	**Percent Exact**	**Percent Exact and Adjacent**	**Inter-rater Reliability**	**Kappa N with Exact Agreement**
Indicator					
1a.1	307	0.78	0.98	0.52	0.71
1a.2	307	0.76	0.99	0.51	0.67
1a.3	307	0.74	0.98	0.45	0.65
1b.1	307	0.61	0.97	0.40	0.47
1b.2	307	0.58	0.97	0.34	0.44
1b.3	307	0.62	0.97	0.40	0.49
1c.1	307	0.67	0.99	0.56	0.56

	N of Items Scored	Rater Agreement		Inter-rater Reliability	Kappa N with Exact Agreement
		Percent Exact	Percent Exact and Adjacent		
1c.2	307	0.59	0.99	0.41	0.45
2a.1	232	0.81	0.99	0.42	0.75
2a.2	232	0.78	0.99	0.49	0.70
2b.1	232	0.82	1.00	0.65	0.76
2b.2	232	0.69	0.99	0.70	0.59
2c.1	232	0.73	0.99	0.66	0.64
2c.2	232	0.75	0.99	0.58	0.66
3a.1	237	0.70	0.98	0.63	0.61
3a.2	237	0.79	1.00	0.73	0.72
3b.1	237	0.80	1.00	0.58	0.74
3b.2	237	0.85	1.00	0.55	0.80
3c.1	237	0.76	0.99	0.39	0.68
3c.2	237	0.72	1.00	0.60	0.63
3c.3	237	0.68	0.98	0.41	0.57
3d.1	237	0.70	1.00	0.58	0.59
4a.1	267	0.77	0.99	0.46	0.70
4a.2	267	0.79	0.99	0.47	0.73
4a.3	267	0.73	0.99	0.55	0.64
4b.1	267	0.67	0.99	0.57	0.57
4c.1	267	0.63	0.99	0.60	0.51
4c.2	267	0.66	1.00	0.60	0.55

In addition to examining rater agreement at exact and adjacent, agreement was also reviewed by indicator as a grouping to see the % 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; this directionality helps to identify indicators where there is more of a disparity between raters. Agreement rates, as shown in Table 13, 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 1c2 has the lowest combination of agreement between both exact agreement and directional agreement.

Table 13. Rater Exact Agreement and Agreement Directionality by Indicator

Agreement Information at the Indicator Level			
		Percent (%)	
Rubric	**Indicator**	**Exact Agreement**	**Both Scorers <2 or >3**
Rubric 1a	1a1	78	86
	1a2	76	82
	1a3	74	81
Rubric 1b	1b1	61	75
	1b2	58	72
	1b3	62	75
Rubric 1c	1c1	67	80
	1c2	59	65
Rubric 2a	2a1	81	91
	2a2	78	90
Rubric 2b	2b1	82	90
	2b2	69	79
Rubric 2c	2c1	73	87
	2c2	75	88
Rubric 3a	3a1	70	88
	3a2	79	90

Agreement Information at the Indicator Level			
		Percent (%)	
Rubric	Indicator	Exact Agreement	Both Scorers <2 or >3
Rubric 3b	3b1	80	89
	3b2	85	90
Rubric 3c	3c1	76	89
	3c2	72	92
	3c3	68	78
Rubric 3d	3d1	70	76
Rubric 4a	4a1	77	80
	4a2	79	84
	4a3	73	79
Rubric 4b	4b1	67	79
Rubric 4c	4c1	63	73
	4c2	66	70

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 % 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 14 indicates the % 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 14. 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	91	4
Scorer 2 Fail	3	3

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	1

Task 3 Decision Consistency Between Scorers at Task Score 2.1 (%)		
	Scorer 1 Pass	Scorer 1 Fail
Scorer 2 Pass	97	1
Scorer 2 Fail	0	1

Task 4 Decision Consistency Between Scorers at Task Score 2.1 (%)		
	Scorer 1 Pass	Scorer 1 Fail
Scorer 2 Pass	90	4
Scorer 2 Fail	3	3

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. Retreived 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	0.98	3.26	0.33
	3	0.33	7.17	71.01	3.58
	2	0.33	6.51	4.89	0.33
	1	0.33	0.65	0.33	0
		1	2	3	4
		Scorer 2			

Joint Score Distribution for Task 1 Indicator 1a2 (%)				
Scorer 1	4	0	0.98	2.61	0.65
	3	0	7.49	69.38	3.26
	2	0	4.89	8.79	0.33
	1	0.65	0.98	0	0
		1	2	3	4
		Scorer 2			

Joint Score Distribution for Task 1 Indicator 1a3 (%)				
Scorer 1	4	0	0.33	2.61	0.65
	3	0	10.1	68.08	3.26
	2	0.33	4.89	6.84	0.98
	1	0.33	0.98	0.65	0
		1	2	3	4
		Scorer 2			

Joint Score Distribution for Task 1 Indicator 1b1 (%)				
Scorer 1	4	0	0.98	5.86	0.98
	3	0	10.1	51.14	6.84
	2	0.33	8.14	11.4	1.95
	1	0.33	1.63	0	0.33
		1	2	3	4
		Scorer 2			

	Joint Score Distribution for Task 1 Indicator 1b2 (%)				
Scorer 1	4	0	0.98	4.56	0
	3	0.65	11.4	51.47	7.49
	2	0.33	5.86	14.01	0.65
	1	0.33	1.63	0.65	0
		1	2	3	4
		Scorer 2			

	Joint Score Distribution for Task 1 Indicator 1b3 (%)				
Scorer 1	4	0	0.65	6.19	0.98
	3	0	11.4	50.49	5.86
	2	0.65	10.1	11.07	1.63
	1	0	0.65	0	0.33
		1	2	3	4
		Scorer 2			

	Joint Score Distribution for Task 1 Indicator 1c1 (%)				
Scorer 1	4	0	0.33	4.23	0.33
	3	0	9.77	59.61	6.84
	2	0.65	5.86	9.12	0.33
	1	1.3	0.98	0.65	0
		1	2	3	4
		Scorer 2			

	Joint Score Distribution for Task 1 Indicator 1c2 (%)				
Scorer 1	4	0	0.33	2.28	0
	3	0	14.98	44.3	2.61
	2	0.98	14.01	18.89	0.33
	1	0.33	0.65	0.33	0
		1	2	3	4
		Scorer 2			

TASK 2

Joint Score Distribution for Task 2 Indicator 2a1 (%)					
Scorer 1	4	0	0.43	3.45	2.16
	3	0	3.45	78.45	6.03
	2	0.43	0.43	4.31	0
	1	0	0.43	0.43	0
		1	2	3	4
		Scorer 2			

Joint Score Distribution for Task 2 Indicator 2a2 (%)					
Scorer 1	4	0	0.43	3.45	0.86
	3	0	4.31	73.28	8.19
	2	0	3.02	5.17	0
	1	0.43	0.43	0.43	0
		1	2	3	4
		Scorer 2			

Joint Score Distribution for Task 2 Indicator 2b1 (%)					
Scorer 1	4	0	0	3.88	2.59
	3	0	5.6	75.86	3.45
	2	0	3.02	4.31	0
	1	0.86	0	0.43	0
		1	2	3	4
		Scorer 2			

Joint Score Distribution for Task 2 Indicator 2b2 (%)					
Scorer 1	4	0	0.43	5.17	3.88
	3	0.86	10.34	49.14	3.88
	2	0.43	15.09	9.05	0
	1	0.86	0.86	0	0
		1	2	3	4
		Scorer 2			

Joint Score Distribution for Task 2 Indicator 2c1 (%)					
Scorer 1	4	0	0	5.6	1.72
	3	0.43	6.47	63.36	6.9
	2	0.86	6.9	5.6	0.43
	1	0.86	0.86	0	0
		1	2	3	4
		Scorer 2			

Joint Score Distribution for Task 2 Indicator 2c2 (%)					
Scorer 1	4		0.86	5.6	1.72
	3		5.6	62.93	7.76
	2		9.91	5.17	0.43
	1				
		1	2	3	4
		Scorer 2			

TASK 3

Joint Score Distribution for Task 3 Indicator 3a1 (%)					
Scorer 1	4	0	0.42	7.59	5.06
	3	0	6.75	59.49	9.28
	2	0.42	5.06	3.38	0.42
	1	0.84	0.42	0.84	0
		1	2	3	4
		Scorer 2			

Joint Score Distribution for Task 3 Indicator 3a2 (%)					
Scorer 1	4	0	0	5.06	3.8
	3	0	4.64	70.89	5.06
	2	0.42	2.95	5.49	0
	1	1.69	0	0	0
		1	2	3	4
		Scorer 2			

Joint Score Distribution for Task 3 Indicator 3b1 (%)					
Scorer 1	4		0	3.8	3.8
	3		3.8	73.42	4.64
	2		2.95	6.75	0.42
	1		0.42	0	0
		1	2	3	4
		Scorer 2			

Joint Score Distribution for Task 3 Indicator 3b2 (%)					
Scorer 1	4		0	2.53	1.27
	3		4.64	80.59	2.53
	2		3.38	5.06	0
	1				
		1	2	3	4
		Scorer 2			

Joint Score Distribution for Task 3 Indicator 3c1 (%)					
Scorer 1	4		0.84	5.06	1.69
	3		5.49	70.89	8.02
	2		3.38	4.64	0
	1				
		1	2	3	4
		Scorer 2			

Joint Score Distribution for Task 3 Indicator 3c2 (%)					
Scorer 1	4	0	0	7.59	4.22
	3	0	5.49	62.45	11.39
	2	0.42	5.49	2.95	0
	1				
		1	2	3	4
		Scorer 2			

Joint Score Distribution for Task 3 Indicator 3c3 (%)					
Scorer 1	4		1.27	3.8	1.27
	3		9.28	57.81	6.33
	2		8.44	11.39	0.42
	1				
		1	2	3	4
		Scorer 2			

Joint Score Distribution for Task 3 Indicator 3d1 (%)					
Scorer 1	4		0	3.8	1.69
	3		12.24	55.27	2.53
	2		12.66	11.81	0
	1				
		1	2	3	4
		Scorer 2			

TASK 4

Joint Score Distribution for Task 4 Indicator 4a1 (%)					
Scorer 1	4	0	0	0.37	0
	3	0.37	9.74	71.16	1.12
	2	0.75	5.99	9.74	0
	1	0	0.37	0.37	0
		1	2	3	4
		Scorer 2			

Joint Score Distribution for Task 4 Indicator 4a2 (%)					
Scorer 1	4	0	0	3	0.37
	3	0.75	7.12	75.66	1.5
	2	0	2.62	8.24	0
	1	0.75	0	0	0
		1	2	3	4
		Scorer 2			

Joint Score Distribution for Task 4 Indicator 4a3 (%)					
Scorer 1	4	0	0	1.5	0
	3	0.37	10.11	66.67	1.12
	2	1.5	6.37	10.11	0
	1	0	1.87	0.37	0
		1	2	3	4
		Scorer 2			

Joint Score Distribution for Task 4 Indicator 4b1 (%)					
Scorer 1	4	0	0	4.49	0.75
	3	0.75	8.99	59.93	4.49
	2	1.12	5.99	11.61	0
	1	0.75	1.12	0	0
		1	2	3	4
		Scorer 2			

Joint Score Distribution for Task 4 Indicator 4c1 (%)					
Scorer 1	4	0	0	3.37	0.75
	3	0	10.49	46.07	4.49
	2	0.37	15.36	16.1	0
	1	0.75	1.5	0.75	0
		1	2	3	4
		Scorer 2			

		Joint Score Distribution for Task 4 Indicator 4c2 (%)			
Scorer 1	4	0	0	0.75	0.37
	3	0	12.36	46.82	2.25
	2	0.75	17.6	16.85	0
	1	1.12	0.75	0.37	0
		1	2	3	4
		Scorer 2			