APPENDIX R

MCAS Validity Evidence

Gathered by the Massachusetts Department of Elementary and Secondary Education

Purpose

Each year, assessment researchers at the Massachusetts Department of Elementary and Secondary Education evaluate validity evidence associated with the MCAS tests. This evidence and the methods for conducting each study and analysis are reviewed by the MCAS TAC for methodological soundness and for appropriateness in responding to the research questions.

In the initial year of the Next-Generation MCAS assessments, the Department conducted a Learning Labs study to better understand educator and student responses to new ELA essay item types in order to provide appropriate practice materials and other supports.

The Department is also engaged in collecting and reporting on evidence that addresses Next-Generation MCAS validity of results with respect to how these results correspond with other variables. The Department is investigating the extent to which these results correspond with the following variables:

- What is the correspondence between students' MCAS results to retention in grade later that year? This relationship provides evidence for the extent to which MCAS signals students' lack of readiness for the next grade level.

MCAS Learning Labs Study

Introduction

The Next-Generation MCAS tests present new item types to students in schools and districts. As some of these item types were based on the new items in PARCC tests, many Massachusetts students who sat for PARCC tests were familiar with these items. Over 1/3 of Massachusetts students, however, did not take PARCC tests in previous years and were unfamiliar with these item types.

One item type in particular, the new ELA essay item that asked students to respond to a single or paired text in an essay format was unfamiliar to a number of Massachusetts students. Previously, on Legacy MCAS tests, students in grades 4, 7, and 10 were askeAppd to provide an essay response in the absence of an ELA passage to respond to. When students were asked to provide an open response to demonstrate reading comprehension on the Legacy MCAS tests, they did so in a short response that did not require the multiple paragraphs or organization of an essay.

The State conducted a series of ELA Learning Labs in the fall of 2016 to better understand how students would respond to the new ELA essay items, as well as to learn from teachers the types of supports that would help educators prepare students for the new item type.

Research Questions

The research questions investigated in the study included:

- Are the new items functioning as intended (and across all student groups)?
    - Are students writing essays in response to the new item type?
    - Are students using evidence from the passages in their answers?
- Are revisions needed to the scoring materials or scoring directions?

Sample

The study was conducted by test developers at the Department of Elementary and Secondary Education (DESE). The study used essay items linked to ELA passages taken in 2016; hence, the new item types

were administered to students in the grade level above (e.g., since students in grade 3 in the spring of 2016 were in grade 4 in the fall, the earliest grade studied was grade 4.
The study was conducted with 37 Teachers in 11 districts, as shown in Table O.1 . Note that in some classrooms more than one teacher responded for a total of 41 teacher participants.

Table R.1: Number of Classrooms in the Learning Labs Study

District Name	G4	G5	G6	G7	G8	G9
Boston			1			
Dracut			1			
East Bridgewater	2					
Franklin					9	
Hopedale	4					1
Longmeadow				7		
Lowell	2					
Palmer		3				
West Springfield		5				
Winchester					2	
Total by Teacher	8	8	2	7	11	1

Districts selected represented a range of district types and demographic categories including: urban and suburban districts, high-, moderate-, and low-achieving districts, high-, moderate-, and low minority ethnic/racial status composition, and varying levels of low income students and English language learners.

Method

The study was conducted by test developers from the DESE. The study was conducted in classrooms with teachers and their students. Prior to the study, DESE developed study protocols and materials that included:

- Test booklets with the new ELA item types. The new item types were revised MCAS items from the previous test administration; consequently, the new item types used in this study could be released as practice materials at the end of the study.
- A study protocol that detailed the test developer and teacher activities during the study, including a script for how to introduce the study to teachers and students, and questions to ask.
- A teacher feedback questionnaire included open-ended questions and a few rating scale questions

After speaking with the teacher and introducing the activities to students, test developers administered the short assessment with the new test items and, along with the classroom teacher, monitored students taking the tests and answered questions from students. After students took the test, test developers and/or the classroom teacher conducted a focus group with students to ask how the session went and areas that were difficult for students. The test developers then interviewed teachers to obtain their feedback on the session, or they asked teachers to fill out the feedback form on their own.

All student booklets were scored by the test development team; during the scoring process the test development team had an opportunity to discuss and revise the scoring materials as needed. These discussions led to refinements in scoring the new item types. Information from the teacher feedback forms was transcribed into an Excel database for analysis.

Findings

Findings from the educator feedback form are provided in Table O.2. The percentages are based on the number of educators queried. Key findings are described below.

- Expectations Clear: Teachers who wrote about this issue in their feedback forms generally indicated that students were unclear about what was expected of them in these essay responses. Two educator comments illustrated the need for the Department to publish practice materials to increase awareness of this new item type:
    - "Students found the new expectations unclear -- students had questions about the length of the response -- some paid more attention to punctuation/spelling and others payed little regard to the instructions." (Grade 7)
    - "The term "essay" will need to be clarified -- what makes an essay different from an open response item? Stress the scoring of mechanics, conventions, and spelling." (Grade 4)
- Took Test Seriously: The majority of teachers noted that students took the tests seriously, indicating that study results were not likely undermined by a lack of effort on behalf of students.
- Test-Taking Strategies Used: About half of the teachers noted that students were using test-taking strategies while taking the assessment. This is another indicator that students were responding to the new item types as expected.
- Took More than One Hour: Almost half of the teachers noted that students took more than 1 hour on the task, which was one indicator of the amount of time students would likely spend on the new assessments.

Table R.2: General Educator Feedback

General Educator Feedback												
	Expectations Clear			Took Test Seriously			Test-Taking Strategies Used			Took More than 1 Hour		
Grade	No	Didn't Mention	Yes	No	Didn't Mention	Yes	No	Didn't Mention	Yes	No	Didn't Mention	Yes
4	67%	33%	0%	0%	20%	80%	29%	0%	71%	0%	80%	20%
5	50%	40%	10%	0%	33%	67%	25%	50%	25%	10%	20%	70%
6	50%	0%	50%	0%	0%	100%	50%	0%	50%	50%	50%	0%
7	60%	40%	0%	17%	67%	17%	67%	33%	0%	20%	20%	60%
8	73%	18%	9%	18%	36%	45%	0%	9%	91%	13%	38%	50%
9	0%	100%	0%	0%	100%	0%	0%	100%	0%	0%	100%	0%
Total %	60%	31%	9%	9%	38%	53%	26%	23%	51%	13%	39%	48%
Total N	21	11	3	3	13	18	9	8	18	4	12	15

The test development team scored the student essays individually. Once the essays were scored, the team discussed two reactions to the student results. First, they reviewed potential changes to the scoring materials for the new essay item types. Second, they planned practice materials and other supports to illustrate the new expectations for teachers and students. The team also developed and delivered an online presentation to address the new item expectations, using results from the study to inform the presentation.

Responses to the new item types were also compared to student responses to the old MCAS items (the MCAS items that the study items were based on) for students in grade 5/6 and grade 7/8 who took MCAS the previous year. Two districts were included in this analysis:

- Dracut (n=20)
- Winchester (n=42)

Correlations between the original item (Item Diff #16) and the revised item type (Item New Diff) were low and insignificant in the grade 5/6 sample (r = .28) but moderate and significant in the grade 7/8 sample indicating that students performance on the item was somewhat different between the two administrations.

Correlations between performance on the new item and the scaled score from the prior year were moderate-to-high and significant for both samples (G5/6 r=.66, G7/8 r=.85), indicating that student performance on the new items was associated with their past performance on the ELA test, as expected.

Table R.3: Learning Lab Results

Correlations: 2017 Learning Lab Results with 2016 MCAS Test Results							
	Item Diff# 16	ItemDev Diff	ItemConv Diff	ItemNew Diff	Escaleds 16	OpenResp 16	Escaleds 14
Grade 5 2016							
Item Diff 16	1						
ItemDev Diff	0.284	1					
ItemConv Diff	0.145	.752**	1				
ItemNew Diff	0.284	.939**	.933**	1			
Escaleds 16	0.395	0.347	.308	0.343	1		
OpenResp 16	.730**	0.301	.184	0.253	.657**	1	
Escaleds 14	0.070	0.074	.205	0.147	.578**	0.374	1
Grade 7 2016							
Item Diff 16	1						
ItemDev Diff	.465**	1					
ItemConv Diff	.385*	.799**	1				
ItemNew Diff	.457**	.969**	.923**	1			
Escaleds 16	.681**	.586**	.571**	.610**	1		
OpenResp 16	.879**	.603**	.535**	.606**	.845**	1	
Escaleds 14	.471**	.468**	.701**	.523**	.665**	.627**	1
**. Correlation is significant at the 0.01 level (2-tailed).							
*. Correlation is significant at the 0.05 level (2-tailed).							
# Indicates difficulty (e.g., ItemDev Difficulty=p-value of the topic development section)							

Next Generation MCAS Validity Evidence: Relationships with Other Variables

Convergent Validity Evidence

Relationships among MCAS results and other variables were evaluated in two analyses. In the first analysis we examined convergent and validity by comparing the magnitude of the correlations of student results (scaled scores) between portions of the assessment within each subject area (convergent validity evidence) and between portions of the assessment outside of each subject area (divergent validity evidence).

The correlation coefficients provided in Table XX.3 show the relationships of the selected response (e.g., ELA SR) and the constructed response (e.g., ELA CR) portions of the test with the scaled scores within each subject area (Escaleds = ELA scaled score) and (Mscaleds = Math scaled score). Evidence of convergent validity would include correlation coefficients within subject that are of a greater magnitude than correlation coefficients outside of subject area. Convergent validity evidence is illustrated with bolded text. A lack of evidence of convergent validity is shown using peach shading. Across the six grades, the same pattern emerges; evidence of convergent validity is shown between:

- ELA: Selected response items and the scaled score
- ELA: Constructed response items and the scaled score
- Math: Selected response items and the constructed response items
- Math: Both the selected and constructed response items and the scaled score.

These results indicate that the test segments featuring the two different item types are both picking up construct relevant information because the segments of each test are more highly correlated with their respective content area than with the alternate content area.

There is a lack of convergent validity evidence in ELA between results on the selected response items and the constructed response items – the range of these correlations is of no greater magnitude than the correlations between the math and ELA selected response results and the math and ELA constructed response result. This suggests that although both segments of the ELA test demonstrates convergence with the overall scaled score, the individual segments appear to require different skill sets (e.g., reading and writing in response to text).

Table R.4: Convergent Validity Evidence Across Two Test Subjects

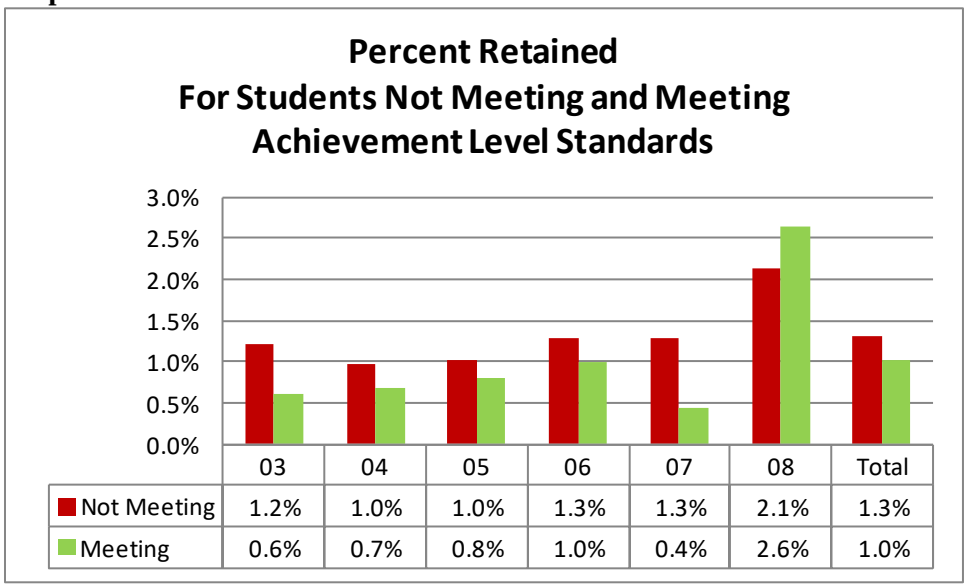
	Grade 3					
	ELA SR	ELA CR	Escaleds	Math SR	Math CR	Mscaleds
ELA SR						
ELA CR	0.63					
Escaleds	**0.93**	**0.84**				
Math SR	0.68	0.60	0.69			
Math CR	0.69	0.61	0.71	**0.85**		
Mscaleds	0.70	0.62	0.74	**0.95**	**0.94**	
	Grade 4					
	ELA SR	ELA CR	Escaleds	Math SR	Math CR	Mscaleds
ELA SR						
ELA CR	0.61					
Escaleds	**0.89**	**0.85**				
Math SR	0.65	0.60	0.68			
Math CR	0.66	0.63	0.71	**0.86**		
Mscaleds	0.67	0.64	0.72	**0.94**	**0.97**	
	Grade 5					
	ELA SR	ELA CR	Escaleds	Math SR	Math CR	Mscaleds
ELA SR						
ELA CR	0.60					
Escaleds	**0.87**	**0.87**				
Math SR	0.63	0.59	0.67			
Math CR	0.64	0.61	0.70	**0.84**		
Mscaleds	0.66	0.62	0.72	**0.95**	**0.95**	
	Grade 6					
	ELA SR	ELA CR	Escaleds	Math SR	Math CR	Mscaleds
ELA SR						
ELA CR	0.66					
Escaleds	**0.91**	**0.88**				
Math SR	0.67	0.62	0.70			
Math CR	0.70	0.63	0.74	**0.84**		
Mscaleds	0.72	0.65	0.75	**0.92**	**0.97**	
	Grade 7					
	ELA SR	ELA CR	Escaleds	Math SR	Math CR	Mscaleds
ELA SR						
ELA CR	0.67					
Escaleds	**0.90**	**0.89**				
Math SR	0.66	0.62	0.69			
Math CR	0.67	0.62	0.72	**0.83**		
Mscaleds	0.70	0.65	0.75	**0.93**	**0.96**	
	Grade 8					
	ELA SR	ELA CR	Escaleds	Math SR	Math CR	Mscaleds
ELA SR						
ELA CR	0.67					
Escaleds	**0.89**	**0.91**				
Math SR	0.68	0.64	0.71			
Math CR	0.64	0.64	0.71	**0.86**		
Mscaleds	0.68	0.66	0.73	**0.96**	**0.95**	

Relationship of MCAS Results to Student Retention

In the second analysis we examined the extent to which the Next-Generation MCAS results signaled a lack of readiness to progress to the next grade. Here it is important to note that because teachers did not have the results of the MCAS assessments when deciding to retain students in grade, the test results had no influence on the retention decision. One additional note on retention – it is a weak criterion because not many students are retained each year (1.2% of the students were retained in grade during the year of this study), and other variables besides academic progress factor into retention, chiefly attendance and student behavior in the classroom/

MCAS proficiency was dichotomized across the two subject areas – students who scored at the "Meeting" or "Exceeding" expectations levels in both subjects were identified as "Meeting" expectations; those who did not meet expectations in either subject or both subjects were identified as "Not Meeting" expectations. The percentages of students retained are shown in Graph O.1 by grade level; those were did not meet the achievement standards are shown in the red bars and those meeting the achievement standards are shown in the green bars. The relationship between retention and meeting expectations is as expected in all grades except for grade 8 wherein students who did not meet achievement standards are retained more often than those who did meet the standards.

Graph R.1



	03	04	05	06	07	08	Total
Not Meeting	1.2%	1.0%	1.0%	1.3%	1.3%	2.1%	1.3%
Meeting	0.6%	0.7%	0.8%	1.0%	0.4%	2.6%	1.0%