

## **Missouri End-of-Course Assessments Comparison of ARC Hand Scoring with Teacher Scores**

### **Introduction**

The Missouri End-of-Course (EOC) Assessments were created to adapt the State's tests needs of Missouri districts, schools, teachers, and students, while still meeting state and federal accountability requirements. The Missouri State Board of Education identified the following purposes for the Missouri EOC Assessments:

- Measuring and reflecting students' mastery toward post-secondary readiness
- Identifying students' strengths and weaknesses
- Communicating expectations for all students
- Serving as the basis for state and national accountability plans
- Evaluating programs

Course Level Expectations (CLEs) outline the ideas, concepts, and skills that form the foundation for an assessed EOC subject area, regardless of student grade level. Because a course such as Algebra I could be delivered in middle school or any grade level in secondary school, CLEs replaced the Grade Level Expectations (GLEs). Districts can offer courses with different titles that cover the same CLEs.

Each Algebra I, English II, and Biology EOC assessment includes two types of test items: selected-response items and performance events (PEs), which include writing prompts. The selected-response items present students with a question followed by four response options. The performance events require students to work through more complicated items. Performance events often allow for more than one approach to arrive at a correct response. The advantage of this type of assessment item is that it provides insight into a student's ability to apply knowledge and understanding in real-life situations.

The writing prompt, a special type of performance event that appears in the English II assessment, is an open-ended item that requires students to demonstrate their writing proficiency. Writing is scored holistically using a four-point scoring guide.

### **Hand-Scoring of PE Items for the Operational EOC Assessments**

Student responses to the operational PE items were hand scored for state reporting and federal AYP determinations by professionally trained raters at the Assessment Resource Center (ARC). ARC provided these raters with extensive training using industry standards for qualification, including formal educational background and percent exact agreement criteria, to determine eligibility for scoring PEs on the EOC Assessments

For the Spring 2009 administration of the Missouri EOC assessments, the following operational PE items were scored:

<u>EOC Assessment</u>	<u>Points Possible</u>
English II Writing Prompt	1 – 4
Algebra I PE Item	0 – 4
Biology PE Items	
PE Item 1	0 – 1
PE Item 2	0 – 1
PE Item 3	0 – 1
PE Item 4	0 – 4
PE Item 5	0 – 2
PE Item 6	0 – 2
PE Item 7	0 – 2
PE Item 8	0 – 1
PE Item 9	0 – 1
PE Item 10	0 – 3
PE Item 11	0 – 2
Total Score	0 – 20

Missouri teachers were also given the opportunity to score their students' responses to the PEs on the operational assessments. Teachers had access to scanned images of their students' response to each PE through an internet-based teacher interface. Teachers were provided online training via a web-based, interactive tutorial designed to help them learn how to score the PEs. That tutorial included the same anchor and training papers that were used by the professional scorers at ARC. Teachers also had the option to use the resulting scores as a component of a student's final course grade.

The purpose of the current study is to provide a comparison of the official ARC PE scores with the scores provided by Missouri teachers. The two scores (i.e., the ARC score and the teacher score) were obtained through a process of matching barcode numbers from student response booklets. The resulting analyses provided in the next section focus on the percent agreement between ARC scores and teacher scores. The implications of the results for Missouri teachers and state policy makers are then considered.

## Results

For each EOC assessment, the ARC score and the teacher score were compared by examining (a) the percent agreement at each score point; (b) the percent exact agreement overall and exact plus adjacent agreement overall; (c) the difference between mean PE scores overall; and (d) the Pearson correlation between the two sets of scores.

## English II

The results for English II are presented in Tables 1 through 3. Table 1 provides a cross-tabulation of ARC scores with Missouri teacher scores. Table 2 provides the frequency distribution of the difference between the teacher scores and the ARC scores. Table 3 provides the mean ARC score and the mean teacher score, including the Pearson correlation between the two sets of scores.

Examination of Tables 1 and 2 shows the overall exact agreement between ARC scores with Missouri teacher scores to be 49.5%. The overall exact plus adjacent agreement was found to be 95.0%. While these percentages are encouraging, especially with respect to the use of EOC student scores for the purpose of assigning course grades, they are lower than the generally accepted standard in the industry. Typically, when professional raters are scoring a constructed-response item with a 0 to 4 scoring rubric, exact agreement is expected to be approximately 80.0% and exact plus adjacent agreement is expected to be 99.0%. Still, given that teachers self-trained using an online tutorial with no supervision or qualifying required, the agreement rates seem reasonable given the intended use of the scores.

**Table 1: English II Cross-Tabulation of ARC Scores with Missouri Teacher Scores**

MO EOC, Spring 2009 PE Scoring ARC Item Score by Teacher Item Score							
Subject	Item	ARC Item Score	Teacher Item Score				
			0	1	2	3	4
English II	1	0	376	12	28	17	1
			1.59	0.05	0.12	0.07	0.00
			86.64	2.76	6.45	3.92	0.23
		1	38	216	92	10	2
			0.16	0.91	0.39	0.04	0.01
			10.61	60.34	25.70	2.79	0.56
		2	52	649	1443	495	45
			0.22	2.74	6.10	2.09	0.19
			1.94	24.18	53.76	18.44	1.68
		3	93	571	4805	7263	3166
			0.39	2.41	20.31	30.69	13.38
			0.58	3.59	30.22	45.68	19.91
		4	11	20	332	1511	2414
			0.05	0.08	1.40	6.39	10.20
			0.26	0.47	7.74	35.24	56.30

**Table 2: English II Distribution of the Difference between Teacher Scores and ARC Scores**

MO EOC, Spring 2009 PE Scoring Teacher Total - ARC Total			
Subject	Teacher Total minus ARC Total	N	Percent
English II	-4	11	0.05
	-3	113	0.48
	-2	955	4.04
	-1	7003	29.60
	0	11712	49.50
	1	3765	15.91
	2	83	0.35
	3	19	0.08
	4	1	0.00

The difference between English II mean scores provided in Table 3 and the frequency distribution in Table 2 shows that ARC raters were more lenient than Missouri teachers. The ARC mean score was 2.98, where the mean score for teachers was found to be 2.76. As indicated by standard deviations of 0.96 and 0.72, respectively, the teacher distribution was slightly more variable than the distribution of ARC's raters. This is consistent with the hypothesis that the more careful training and monitoring of professional scorers can lead to more accuracy and precision in scoring. Finally, the correlation between the two sets of scores was .58, which represents a moderate to moderately-high correlation coefficient.

**Table 3: English II Summary Statistics for ARC Scores and Teacher Scores**

MO EOC, Spring 2009 PE Scoring Summary Statistics					
Subject	Scorers	N	Mean	SD	Correlation
English II	ARC	23662	2.98	.72	.58
	Teachers	23662	2.76	.96	

### Algebra I

The results for Algebra I are presented in Tables 4 through 6. Table 4 provides a cross-tabulation of ARC scores with Missouri teacher scores. Table 5 provides the frequency distribution of the difference between the teacher scores and the ARC scores. Table 6 provides the mean ARC score and the mean teacher score, including the Pearson correlation between the two sets of scores. Examination of Tables 4 and 5 shows the

overall exact agreement between ARC scores with Missouri teacher scores to be 49.1%. The overall exact plus adjacent agreement was 84.5%. Similar to the English II results, these percentages seem reasonable, especially with respect to the intended use of the EOC teacher scores.

**Table 4: Algebra I Cross-Tabulation of ARC Scores with Missouri Teacher Scores**

MO EOC, Spring 2009 PE Scoring ARC Item Score by Teacher Item Score							
Subject	Item	ARC Item Score	Teacher Item Score				
			0	1	2	3	4
Algebra I	1	0	1177	481	37	8	2
			5.09	2.08	0.16	0.03	0.01
			69.03	28.21	2.17	0.47	0.12
		1	758	3175	2392	939	300
			3.28	13.74	10.35	4.06	1.30
			10.02	41.98	31.62	12.41	3.97
		2	10	409	3321	2678	1896
			0.04	1.77	14.37	11.59	8.20
			0.12	4.92	39.94	32.21	22.80
		3	2	33	336	840	444
			0.01	0.14	1.45	3.63	1.92
			0.12	1.99	20.30	50.76	26.83
		4	2	15	343	689	2825
			0.01	0.06	1.48	2.98	12.22
			0.05	0.39	8.85	17.79	72.92

The difference between the Algebra I mean scores provided in Table 6 and the frequency distribution in Table 5 shows that unlike the scoring of the English II writing task, Missouri teachers were more lenient than ARC raters when they scored the Algebra I PE. In Algebra I, the ARC mean score was 1.93 where the mean score for teachers was found to be 2.35. The correlation between the two sets of scores was .70 which represents a moderately-high correlation coefficient. This would be consistent with the possibility that the scoring guide for Algebra I might have been less subjective (less call for rater judgment) than the scoring guide for English II.

**Table 5: Algebra I Distribution of the Difference between Teacher Scores and ARC Scores**

MO EOC, Spring 2009 PE Scoring Teacher Total - ARC Total			
Subject	Teacher Total- ARC Total	N	Percent
Algebra I	-4	2	0.01
	-3	17	0.07
	-2	386	1.67
	-1	2192	9.48
	0	11338	49.06
	1	5995	25.94
	2	2872	12.43
	3	308	1.33
4	2	0.01	

**Table 6: Algebra I Summary Statistics for ARC Scores and Teacher Scores**

MO EOC, Spring 2009 PE Scoring Summary Statistics					
Subject	Scorers	N	Mean	SD	Correlation
Algebra I	ARC	23112	1.93	1.17	.70
	Teachers	23112	2.35	1.25	

### Biology

The results for Biology are presented in Tables 7 through 10. Because there are multiple parts or items for Biology, the results for this content area are presented differently than the other two content areas. Table 7 provides the frequency distribution of the difference between the teacher item scores and the ARC item scores. Table 8 provides the frequency distribution of the difference between the teacher total score and the ARC total score. Recall that the total Biology PE score range is from 0 to 20. Table 9 provides a cross-tabulation of ARC scores with Missouri teacher scores for each of the 11 Biology PE items. Table 10 provides the total mean ARC score and the total mean teacher score, including the Pearson correlation between the two sets of scores.

Combining the exact agreement percentage from across all 11 PE items (Table 7), the overall exact agreement between ARC scores with Missouri teacher scores is 66.6%. The overall exact plus adjacent agreement is 93.4%. These percentages are higher than the other content areas because most of the Biology PE items are scored 0 to 1, 0 to 2, or 0 to

3. Similar to the other content areas, these percentages seem reasonable, especially with respect to the intended use of the EOC teacher scores.

**Table 7: Biology Distribution of the Difference between Teacher Scores and ARC Scores for PEs**

<b>MO EOC, Spring 2009 PE Scoring Teacher Item Score - ARC Item Score</b>			
<b>Subject</b>	<b>Teacher Score- ARC Score</b>	<b>N</b>	<b>Percent</b>
<b>Biology</b>	-4	4	0.00
	-3	75	0.03
	-2	1429	0.57
	-1	13433	5.33
	0	167911	66.63
	1	53961	21.41
	2	13442	5.33
	3	1746	0.69
	4	21	0.01

**Table 8: Biology Distribution of the Difference between Teacher Scores and ARC Scores for the Total PE Score**

<b>MO EOC, Spring 2009 PE Total Scoring Teacher Total - ARC Total</b>			
<b>Subject</b>	<b>Teacher Total- ARC Total</b>	<b>N</b>	<b>Percent</b>
<b>Biology</b>	-11	1	0.00
	-10	3	0.01
	-9	28	0.12
	-8	22	0.09
	-7	40	0.17
	-6	65	0.28
	-5	68	0.29
	-4	113	0.48
	-3	229	0.97
	-2	527	2.24
	-1	967	4.11
	0	2102	8.93
	1	2786	11.84
	2	3518	14.95
	3	3701	15.73
	4	3201	13.60
	5	2473	10.51
	6	1603	6.81
	7	988	4.20
	8	532	2.26
	9	272	1.16
10	157	0.67	
11	69	0.29	
12	35	0.15	
13	15	0.06	
14	5	0.02	
15	5	0.02	
16	3	0.01	
17	2	0.01	
18	1	0.00	
20	1	0.00	

**Table 9: Biology Cross-Tabulation of ARC Scores with Missouri Teacher Scores**

MO EOC, Spring 2009 PE Scoring ARC Item Score by Teacher Item Score							
Subject	Item	ARC Item Score	Teacher Item Score				
			0	1	2	3	4
Biology	1	0	3442 14.86 47.20	3850 16.62 52.80			
		1	856 3.69 5.39	15020 64.83 94.61			
	2	0	4139 17.86 91.49	385 1.66 8.51			
		1	123 0.53 0.66	18527 79.95 99.34			
	3	0	5081 21.94 87.36	735 3.17 12.64			
		1	134 0.58 0.77	17210 74.31 99.23			
	4	0	473 2.03 43.51	313 1.35 28.79	187 0.80 17.20	93 0.40 8.56	21 0.09 1.93
			256 1.10 9.97	830 3.57 32.33	776 3.34 30.23	505 2.17 19.67	200 0.86 7.79
		2	143 0.61 2.82	641 2.75 12.66	1727 7.42 34.11	1753 7.53 34.62	799 3.43 15.78
			27 0.12 0.34	287 1.23 3.61	1008 4.33 12.67	3792 16.30 47.67	2841 12.21 35.71
		4	4 0.02 0.06	19 0.08 0.29	113 0.49 1.71	1080 4.64 16.37	5380 23.12 81.56
	5	0	2440 10.53 24.74	3221 13.90 32.66	4200 18.12 42.59		
			321 1.39 3.44	2427 10.47 26.00	6587 28.42 70.56		
		2	80 0.35 2.01	660 2.85 16.58	3240 13.98 81.41		

MO EOC, Spring 2009 PE Scoring ARC Item Score by Teacher Item Score							
Subject	Item	ARC Item Score	Teacher Item Score				
			0	1	2	3	4
Biology	6	0	2067	2536	2336		
			9.12	11.19	10.30		
			29.79	36.55	33.66		
	6	1	671	3071	5603		
			2.96	13.55	24.72		
			7.18	32.86	59.96		
	6	2	279	1252	4854		
			1.23	5.52	21.41		
			4.37	19.61	76.02		
	7	0	2571	1757	969		
			11.28	7.71	4.25		
			48.54	33.17	18.29		
	7	1	385	5222	2415		
			1.69	22.91	10.60		
			4.80	65.10	30.10		
	7	2	103	1231	8136		
			0.45	5.40	35.70		
			1.09	13.00	85.91		
	8	0	4219	1156			
			18.41	5.04			
			78.49	21.51			
8	1	619	16925				
		2.70	73.85				
		3.53	96.47				
9	0	4246	3104				
		18.59	13.59				
		57.77	42.23				
9	1	953	14541				
		4.17	63.65				
		6.15	93.85				
10	0	3046	3861	3193	1453		
		13.49	17.10	14.15	6.44		
		26.37	33.42	27.64	12.58		
		131	468	635	353		
10	1	0.58	2.07	2.81	1.56		
		8.25	29.49	40.01	22.24		
		137	988	2404	3003		
10	2	0.61	4.38	10.65	13.30		
		2.10	15.13	36.80	45.97		
		29	223	807	1842		
10	3	0.13	0.99	3.58	8.16		
		1.00	7.69	27.82	63.50		
		2901	2480	900			
11	0	13.02	11.13	4.04			
		46.19	39.48	14.33			

MO EOC, Spring 2009 PE Scoring ARC Item Score by Teacher Item Score							
Subject	Item	ARC Item Score	Teacher Item Score				
			0	1	2	3	4
Biology	1		659	4911	6950		
			2.96	22.04	31.19		
			5.26	39.23	55.51		
	2		64	658	2759		
			0.29	2.95	12.38		
			1.84	18.90	79.26		

**Table 10: Biology Summary Statistics for ARC Scores and Teacher Scores**

MO EOC, Spring 2009 PE Scoring Summary Statistics					
Subject	Scorers	N	Mean	SD	Correlation
Biology	ARC	23532	11.01	4.21	.79
	Teachers	23532	13.89	4.43	

The difference between the Biology total mean scores provided in Table 10 shows that Missouri teachers were more lenient than ARC raters. The ARC total mean score was 11.01, and the teacher mean score was 13.89. The correlation between the two sets of scores was .79 which represents a relatively high correlation coefficient.

### Possible Implications of the Results

The following bullet points outline some possible implications of the results of the comparison of scoring of PEs between ARC raters and Missouri teachers.

- While teachers in each content area used the teacher interface to score over 23,000 student responses to PE items, the responses for approximately 43,000 other students per content area were not scored by teachers. Because not all students in the state had their responses scored by their teacher, and not all Missouri teachers used the teacher interface for scoring, the sample of teachers who scored items and were compared to ARC raters in the present study could be biased and not representative of the entire population of Missouri teachers. It is conceivable that the teachers who did participate were more convinced of the importance of using the scores as part of course grades, more comfortable with the task of scoring by rubric, and otherwise more engaged than other Missouri teachers.
- While the percentages for exact agreement and exact plus adjacent agreement were reasonable, especially with respect to the current use of EOC student scores

for the purpose of assigning course grades, they were found to be lower than the generally accepted standard in the industry. DESE may want to further evaluate the online training for teachers to determine possible improvements or other activities that could enhance teacher scoring.

- The correlations between the two sets of scores for Algebra I and Biology were moderately high to high, suggesting a strong relationship between how ARC raters approached scoring and how Missouri teachers approached the same scoring task. The higher correlation for Biology could be due in part to the more structured scoring of the items and the larger number of score points. The English II correlation was substantively lower which could reflect differences in the clarity of the scoring rubric, or possible issues with the amount of teacher training required for a writing sample as opposed to the PE items in Algebra I or Biology.

## Summary

In summary, we believe these results to be very encouraging. While the results do not suggest that teachers, who are trained remotely and score unmonitored, can provide an acceptable level of scoring to replace professional scorers, the results do suggest that the teachers in this initial study, on the whole, were very conscientious and able to provide results that were valid for local use.