NISS

Evaluation of Performance-based and Task-based Assessments of Online Research and Comprehension

Analysis of ORCA II Data

Nell Sedransk, Eli Bruner, Weiwei Cui

Technical Report 191
April 2014

National Institute of Statistical Sciences
19 T.W. Alexander Drive
PO Box 14006
Research Triangle Park, NC 27709
www.niss.org

Technical Report

Evaluation of Performance-based and Task-based Assessments of Online Research and Comprehension

Analysis of ORCA II Data

N. Sedransk, E. Bruner, W. Cui
National Institute of Statistical Sciences

9 April 2014

ORCA Project

Don Leu, Project Leader
Jonna M. Kulikowich, Julio Coiro, Nell Sedransk Co-Principal Investigators
University of Connecticut ORCA Team Members

Analysis of Data from ORCA II - Executive Summary

ORCA II is the second in a series of large studies to evaluate a performance-based and a comparable task-based assessment of internet literacy. Over 1800 7th graders at 41 schools in two states participated; of these 1281 completed all background information, an offline reading measure and two assessments, one performance-based and the other task-based. An internet survey was also completed by 1223 students. The performance based assessment was developed as an interactive test with access restricted to an extensive synthetic internet (ORCA-Closed). The task-based assessment was a multiple choice (MC) test administered on a computer screen but following a traditional approach with items that mimicked the items in the performance-based assessment.

The analysis of data from ORCA II addresses eight questions.

1. ***1:*** *How do Multiple Choice and ORCA-Closed correlate with a pure reading measure (Offline Reading Measure)?*
2. ***2:*** *What is the relative difficulty of each component sub-score for ORCA-Closed and what is the relative difficulty for Multiple Choice?*
3. ***3:*** *Do multiple versions give equivalent results for Multiple? Do they for ORCA-Closed? If not how do they differ for each format?*
4. ***4:*** *What is the dimensionality of Multiple Choice? What is the dimensionality of ORCA-Closed?*
5. ***5:*** *How do students perform differently on Multiple Choice and ORCA-Closed? Are there any specific differences evident for particular components?*
6. ***6:*** *How do Multiple Choice and ORCA-Closed scores differ for each of the components and what relationships are evident among component scores?*
7. ***7:*** *What factors drive students' performance, considering Multiple Choice and ORCA-Closed separately, from among test attributes, school attributes, student attributes and student computer familiarity?*
8. ***8:*** *After accounting from general reading ability (through Offline Reading Measure) and version of test, which factors drive students' performance from among test attributes, school attributes, student attributes and student computer familiarity?*
    - ***a:*** *What is the impact of economic differences on students' performance, especially on ORCA-Closed?*
    - ***b:*** *Is the impact of factors such as a socio-economic indicator or computer familiarity different between states – one relatively rich, the other relatively poor; one with a one-to-one laptop policy, the other with no such policy?*

Detailed responses to these eight questions are presented in this technical report with supporting analyses, tables and graphs. Several summary observations transcend the individual responses.

First, the performance based (ORCA-Closed) assessment is multidimensional, that is it draws on multiple distinct skills, whereas the task-based (Multiple Choice) assessment is clearly unidimensional and draws on a single trait.

The design of ORCA-Closed was predicated on a model of online research and comprehension with four components: Locate, Evaluate, Synthesize and Communicate (L,E,S,C). Analysis of the multidimensionality confirmed these four components as not fully independent but nonetheless identified with four distinct skills.

Second, the predominant factor in models of performance on these assessments of internet research and comprehension is baseline reading skill, measured as part of ORCA II by an Offline Reading Measure (ORM). Of the four components, Locate was distinguished from the others because ORM was not a significant factor for the Locate score.

In all cases and all versions, the task-based assessment was on average easier than the performance-based assessment with a difference of from 2.59 to 3.97 points (total points possible = 16) depending on version. Not every student found the task-based assessment easier, almost 10% scored higher on the performance-based assessment.

Gender appeared at first glance to be a highly significant factor with females consistently outscoring males. However, females' Offline Reading Measure scores were also consistently higher than males' scores, thus confounding the comparison. When comparisons took into account ORM, this disappeared and in some cases males outperformed females, especially on the performance-based assessment.

Finally, familiarity with computers, especially outside school, contributed to higher scores on ORCA-Closed but not on the Multiple Choice test.

Introduction: Purpose of study, Design and Participating Students

This is the second in a series of large studies of performance-based and comparable task-based assessments of internet literacy. The first of these, the Online Reading Comprehension Assessment (ORCA I) was administered to over 1300 7th graders in three formats. The performance based assessments were conducted in two formats with identical items. The difference was access to the open internet (ORCA-Open) and access restricted to an extensive synthetic internet (ORCA-Closed). The task-based assessment was a multiple choice (MC) test administered on a computer screen but following a traditional approach with items that mimicked the items in the performance-based assessment. Each student took two different assessments in the same format approximately one week apart.

The Participants in the Study

The second of these, the Online Research and Comprehension Assessment (ORCA II) was administered during the academic year, Fall 2012 through Spring 2013 to over 1800 7th graders. It focused on comparison between ORCA-Closed and MC, as is documented in this Technical Report. In ORCA II each student took two assessments approximately one week apart; however, one was a performance-based assessment (ORCA-Closed) and the other was a traditional test (MC). As was true of ORCA I, students came from pairs of classrooms in schools from two states – one a state with one-to-one laptops in the classroom; the other a state without laptops. Schools were chosen from all socioeconomic levels, all sizes and all performance levels according to their states' official classification or average state exam scores. Table A.1 provides a description of the schools participating in this project.

The Study Design

Four versions of the assessments selected from those tested in ORCA I followed four different scenarios; two versions required a final response in an email format, the other two required a final response in the form of a wiki entry. All possible combinations of version pairs that met the constraint of one ORCA-Closed, one MC and one email response and one wiki were assigned to different students in both possible orders. So on the first testing day in the two classrooms in a single school, only two of the four versions were assigned (1 wiki response, 1 email response). Within each classroom, each of the four Version x Format combinations was assigned at random to approximately one quarter of the students. Different combinations of versions were assigned to different schools for the first day of testing. The second day the alternate versions were assigned in each school, and again all four Version x Format combinations were assigned to some of the students. This was done in carefully planned fashion so that over the course of two testing days each student completed one ORCA-Closed assessment and one Multiple Choice assessment, and these two assessments also comprised one version with wiki response and one with email response.

In addition background information was collected about the school and about each student; and each student completed a specially designed baseline Offline Reading Measure (ORM) and a survey about personal internet use.

Table A.2 provides information about the students who participated; and Table A.3 provides data on the specific forms of the assessments that were administered together with samples sizes for

students completing all or part of the pair of assessments. Analyses in this technical report primarily use 1281 students who took both a Multiple Choice and an ORCA-Closed assessment, or the reduced set of 1223 who also completed the Student Internet Survey.

The Assessments

The assessment paradigm was constructed in accordance with the new literacy theory, treating online reading comprehension as a problem-solving process with four major cognitive components. L, E, S, C.

- **L**ocating information online
- **E**valuating information critically
- **S**ynthesizing information from multiple sources
- **C**ommunicating information also using internet modes.

Thus four items or score points were designed deliberately to measure each of the four components, yielding a 16-point scale with four subscales of 4 points each. Four different scenarios (versions0 were based on different research question posed. If the four components are distinct, then the assessment must draw on multiple latent traits. In such a case the students' ability profiles across this multidimensionality would be expected to differ; this would, in fact, confirm results from the analysis of data from ORCA I.

The Research Questions

A set of eight research questions were posed about the performance-based and the task-based assessments of internet literacy. Analyses responding to these questions make up the remainder of the report with each question; each section is devoted to one question, the response and the supporting analyses.

1: *How do Multiple Choice and ORCA-Closed correlate with a pure reading measure (Offline Reading Measure)?*

2: *What is the relative difficulty of each component sub-score for ORCA-Closed and what is the relative difficulty for Multiple Choice?*

3: *Do multiple versions give equivalent results for Multiple? Do they for ORCA-Closed? If not how do they differ for each format?*

4: *What is the dimensionality of Multiple Choice? What is the dimensionality of ORCA-Closed?*

5: *How do students perform differently on Multiple Choice and ORCA-Closed? Are there any specific differences evident for particular components?*

6: *How do Multiple Choice and ORCA-Closed scores differ for each of the components and what relationships are evident among component scores?*

7: *What factors drive students' performance, considering Multiple Choice and ORCA-Closed separately, from among test attributes, school attributes, student attributes and student computer familiarity?*

8: *After accounting from general reading ability (through Offline Reading Measure) and version of test, which factors drive students' performance from among test attributes, school attributes, student attributes and student computer familiarity?*

   a. *What is the impact of economic differences on students' performance, especially on ORCA-Closed?*

b. *Is the impact of factors such as a socio-economic indicator or computer familiarity different between states – one relatively rich, the other relatively poor; one with a one-to-one laptop policy, the other with no such policy?*

Table A.1. 1820 Students in 2 States from 41 Schools

		State 1	**State 2**
State Level	Median Household Income	$49,158	$64,247
	National Rank	32	4
	1 to 1 Laptop in classroom	No	Yes
	Number of Schools participating	17	24
	Performance Measure	State classification	Mean(reading +math scores)
School Level	Performance Level	Every level A-I	Every decile
	7th Grader Enrollment	12 - 308	55 - 402-
	% free/reduced price lunch	1.6% - 95%	4.8% - 70.9%

Table A.2. Description of Participating Students represented in the ORCAII data base

Student Level	Numbers of Students	Participating Students	≥1 LESC Assessment	2 LESC Assessments	Offline Reading Measure	Internet Use Student Survey
State 1	All students	989	917	827	965	941
	Boys	499*	473	427	493*	439**
	Girls	468*	444	400	452*	482**
State 2	All students	831	784	693	816	803
	Boys	407***	392	344	399***	393****
	Girls	403***	392	349	401***	391****
TOTAL		1820	1701	1520	1781	1744
Format	Test taken	1820	1701	1520	1781	1744
	MC First	896#	896	725	874#	770##
	MC Last	805#	805	795	795#	859##

*: Gender missing for 22 students for state 1.
**: Gender missing for 20 students for state 1.
***: Gender missing for 21 students for state 2.
****: Gender missing for 19 students for state 2.
#: Format information missing for 119 students.
##: Format information missing for 115 students.

Table A.3a Sample Sizes for Assessment Versions for each Format

Student Level		Multiple Choice			ORCA-Closed		
	Version	Participating Students	≥1 LESC Assessment	2 LESC Assessment	Participating Students	≥1 LESC Assessment	2 LESC Assessment
Topic	1-Energy Drinks	417	417	395	364	364	350
	3-Video Games	369	369	343	433	433	416
	5-Heart Healthy Snacks	398	398	370	429	429	407
	7-Contact Lenses	437	437	412	374	374	347
Did not complete a LESC		80	80	--	101	101	--
No LESC taken		119	--	--	119	--	--
TOTAL		1820	1701	1520	1820	1701	1520

Table A.3b Sample Sizes for each Assessment Version for each Format

		Multiple Choice			ORCA-Closed		
Student Level	Version	Participating Students	2 LESCs & Offline Reading Assessment	2 LESCs & ORM & Internet Survey	Participating Students	2 LESCs & Offline Reading Assessment	2 LESCs & ORM & Internet Survey
Topic	1-Energy Drinks	417	361	348	364	271	348
	3-Video Games	369	321	306	433	301	306
	5-Heart Healthy Snacks	398	281	269	429	368	269
	7-Contact Lenses	437	318	300	374	341	300
Did not complete a LESC		80			101		
No LESC taken		119			119		
Total		1820	1281	1223	1820	1281	1223

1: *Question: How do Multiple Choice and ORCA correlate with a pure reading measure (ORM)?*

We investigated the correlations between ORCA and offline reading, and between Multiple Choice and Offline Reading Measure (ORM) as an indicator of the degree to which the performance-based test goes beyond basic reading skills (see *TABLE 1.1*). The correlation between the TOTAL score (Multiple Choice + ORCA ORCA-Closed) and ORM score is 0.56, a degree of correlation that corresponds to a coefficient of determination of 0.31 . The correlations for males (.60) and for females (.54) were similar and both indicate a relationship between online and offline reading as measured by the ORM and ORCA. The score differentials between Multiple Choice and ORCA-Closed were not correlated either with ORM or with the TOTAL score. This was true for males and females, both separately and combined. In other words, whether considering the whole group of students or either gender alone, the relative ease of ORCA-Closed compared to Multiple Choice was not related to the student's overall ability or to the student's reading ability baseline. Of most interest, the correlation between the score on ORCA-Closed and Multiple Choice was only .54 overall, with .56 for males and .50 for females, indicating important differences in what the two assessments measure.

Table 1.1 Correlations between Multiple Choice and ORCA-Closed, also TOTAL and Difference

CORRELATIONS	Offline Reading Score	ORCA Closed Score	Multiple Choice Score	TOTAL Score Closed+MC	Differential Closed -MC	Gender
Sample Size	1281	1281	1281	1281	1281	All
Offline Reading	**1**	0.460	0.531	0.565	-0.081	All
ORCA-Closed	0.460	**1**	0.541	0.876	0.468	All
Multiple Choice	0.531	0.541	**1**	0.880	-0.490	All
TOTAL Score	0.565	0.876	0.880	**1**	-0.017	All
Differential	-0.081	0.468	-0.490	-0.07	**1**	All
Sample Size	612	612	612	612	612	Male
Offline Reading	**1**	0.490	0.563	0.596	-0.100	Male
ORCA-Closed	0.490	**1**	0.564	0.879	0.433	Male
Multiple Choice	0.563	0.564	**1**	0.889	-0.499	Male
TOTAL Score	0.596	0.879	0.889	**1**	-0.048	Male
Differential	-0.100	0.433	-0.499	-0.048	**1**	Male
Sample Size	669	669	669	669	669	Female
Offline Reading	**1**	0.443	0.501	0.545	-0.065	Female
ORCA-Closed	0.443	**1**	0.502	0.864	0.487	Female
Multiple Choice	0.502	0.502	**1**	0.869	-0.511	Female
TOTAL Score	0.545	0.864	0.869	**1**	-0.018	Female
Differential	-0.065	0.487	-0.511	-0.018	**1**	Female

2: *Question: What is the relative difficulty of each component sub-score (separately for ORCA and Multiple Choice)?*

Each of the four LESC components was evaluated for both Multiple Choice and ORCA-Closed to determine the relative difficulty of each component for each format; and to assess the comparative difficulty of each component between formats. For the ORCA-Closed, the range of scores ran from the greatest difficulty with a score of 1.03 (Communicate) to easiest with a score of 2.42 (Synthesis) as shown in Table 2.1. The four disparate LESC scores indicated significant differences between difficulties of each process for the ORCA-Closed format.

With Multiple Choice, difficulty was relatively uniform across components, with Evaluate being the easiest at 2.71 points – a slight but statistically significant disparity - compared to the other three components with average scores of 2.44 to 2.58. The within-student differences between ORCA-Closed and Multiple Choice scores measure the relative difficulty of each component presented in the two formats. For all four components there was a statistically significant difference, ORCA-Closed being more difficult, but with minimal effect size for Locate and Synthesis. However, for Evaluate and for Communicate the highly significant differences were important and constituted 77.5% of the total difference between scores for the two formats.

Table 2.1 Descriptive Statistics for Component Subscores, for TOTAL and for Difference

		Mean	Std Dev	Sample Size	Std Err	Min	Quartile 1	Median	Quartile 3	Max
Closed	LESC	6.80	3.112	1281	0.087	0	5	7	9	15
	Locate	1.83	1.233	1281	0.034	0	1	2	3	4
	Evaluate	1.51	0.975	1281	0.027	0	1	1	2	4
	Synthesis	2.42	1.356	1281	0.038	0	1	3	4	4
	Communicate	1.03	1.065	1281	0.030	0	0	1	2	4
MC	LESC	10.23	3.157	1281	0.088	1	8	11	13	16
	Locate	2.44	1.107	1281	0.031	0	2	3	3	4
	Evaluate	2.71	1.141	1281	0.032	0	2	3	4	4
	Synthesis	2.59	1.140	1281	0.032	0	2	3	4	4
	Communicate	2.49	1.124	1281	0.031	0	2	3	3	4
TOTAL		17.03	5.503	1281	0.154	2	13	17	21	30
Difference	LESC	-3.43	3.004	1281	0.084	-12	-5	-3	-1	5
	Locate	-0.60	1.469	1281	0.041	-4	-2	-1	0	4
	Evaluate	-1.20	1.249	1281	0.035	-4	-2	-1	0	3
	Synthesis	-0.16	1.488	1281	0.042	-4	-1	0	1	4
	Communicate	-1.46	1.321	1281	0.037	-4	-2	-1	-1	3

3: *Question: Do multiple versions give equivalent results for Multiple Choice? Do they for ORCA? If not, how do the results differ for each of Multiple Choice and ORCA?*

The importance of the version of the test was considered separately for ORCA-Closed and for Multiple Choice based on average scores as shown in Table 3.1. Overall, versions differed statistically significantly. For ORCA-Closed versions 1 and 3 were significantly easier; versions 5 and 7 were equivalent to each other at a higher level of difficulty. Both the easier and the harder pair included one version with wiki response and one with email response on the Communicate component. Differences for Multiple Choice were slight with only version 1- Energy Drinks being more difficult than the other versions,

These version differences persisted for ORCA-Closed when results were separated for each gender or for each laptop condition (state), as Tables 3.2 and 3.3 show. For Multiple Choice differences were in the same direction although the results separated by gender were less dramatic.

Since students took one ORCA-Closed and one Multiple Choice assessment, version difference would be particularly important when considering the within-student score difference between the two formats. Consequently, scores for version pairs in Table 3.4 were also analyzed. From examining scores for version pairs, both TOTAL score and Difference (ORCA-Closed score – Multiple Choice score) topic order differentials were investigated. Also the overall score distribution were plotted (Graphic 3.5: Graphs of Score Distributions with Tables) to evaluate differences in the shapes of the distributions between the two formats for each topic pair. Topic orders are noted as topics A-B, with A being the first topic tested.

For topic pairs, the differences between ORCA-Closed and Multiple Choice were compared by means (standard errors were required to test for statistical significance). For 1-3, students scored higher on Multiple Choice than ORCA-Closed by 3 points (10.1 to 7.1); and for 3-1 students scored higher on Multiple Choice by 2.7 points than ORCA-Closed (9.8 to 7.1). Topic order 1-7 noted students scoring 11.0 on Multiple Choice and 7.4 on ORCA-Closed for a difference of 3.; and 7-1 had a differential of 3.3 points (Multiple Choice = 9.9, ORCA-Closed = 6.6. For the case of 5-3, there was a differential of means of 3.6, with Multiple Choice at 10.5 points and ORCA-Closed at 6.9; and 3-5 had a score differential of 4.1 (Multiple Choice = 11.1, ORCA-Closed = 7.0. Lastly, topics 5 and 7 were looking at, with 5-7 having students scoring 9.9 points of Multiple Choice and 6.0 on ORCA-Closed (3.9 point difference); and 7-5 having Multiple Choice scores at 9.8 and students using ORCA-Closed scoring 6.5 (3.3 point difference). Uniformly across the pairings, students scored higher on Multiple Choice – by a minimum of 2.7 points and a maximum of 4.1, indicating offline reading skills to be on the whole easier for students to utilize than Internet skills.

Topic pairs were investigated to determine the importance of order for all topic pairs. Looking first at topics 1 and 3, there were no significant differences between the topics for either Closed or Multiple Choice. For both, the mean scores were within .4 points of each other, with median scores within 1 point of each other on an overall 0-16 scale. For topics 1 and 7, similarly there were no significant differences. Each Closed-Closed and Multiple Choice-Multiple Choice difference of means was less than 1.1 points, with medians again 1 point of lower difference. Topics 5 and 3 had the same general result: means were within .6 points, and medians within 1 point or equal to each

other. Lastly, topics 5 and 7 were within .5 mean points of each other, with median difference of 1 point or equivalence. In sum, there were no significant differences between topic orders across all topic pairings.

Finally we considered the overall shapes of the score distribution for each of the combined scores for Multiple Choice and ORCA-Closed paired versions. On average across the Multiple Choice distributions, the average of median scores was 10.56, with the median of the medians at 10.25. This difference is not surprising in view of the consistent low-end tail for all distributions. This could be a second distribution, i.e., representing a distinct group of low-scoring students, while the rest of the scores follow a relatively normal distribution. Thus the test appears to be well calibrated with the center of the range of scores towards the higher end of the 0-16 point range but without a ceiling effect. The ORCA-Closed distributions had an average median of 7.125, and a median of medians of 7. These distributions were generally followed a normal distribution, and did not show either a floor or ceiling effect since the range of scores was 0-14.

What these distributions clearly show is that students scored higher on Multiple Choice than ORCA-Closed with many fewer students scoring on the low end of the spectrum for Multiple Choice. Thus there is a difference in distributions between formats. Of particular potential interest is a group of low-scoring students that emerges based on scores on the Multiple Choice test.

TABLE 3.1 Scores by Version

	Version	Mean	Std Dev	Sample Size	Std Err	Min	Quartile 1	Median	Quartile 3	Max
Closed	1	7.26	2.90	271	0.18	0	5	7	10	14
	3	7.09	3.17	301	0.18	0	5	7	9	15
	5	6.47	3.18	368	0.17	0	4	6	9	14
	7	6.54	3.10	341	0.17	0	4	7	9	14
MC	1	9.85	3.32	361	0.17	1	7	10	13	16
	3	10.34	2.98	321	0.17	2	8	11	13	16
	5	10.44	3.08	281	0.18	3	8	11	13	16
	7	10.36	3.20	318	0.18	1	8	11	13	16

TABLE 3.2 Scores by Version separately by Gender

	Version	Gender	Mean	Std Dev	Sample Size	Std Err	Min	Quartile 1	Median	Quartile 3	Max
ORCA-Closed	1	Male	6.89	3.00	133	0.26	0	5	7	9	14
	1	Female	7.62	2.76	138	0.23	1	6	7	10	14
	3	Male	6.40	3.27	145	0.27	1	4	6	8	14
	3	Female	7.74	2.93	156	0.23	0	6	8	10	15
	5	Male	5.82	3.11	173	0.24	0	3	6	8	14
	5	Female	7.05	3.14	195	0.22	0	5	7	9	14
	7	Male	5.78	3.19	161	0.25	0	3	6	8	13
	7	Female	7.22	2.85	180	0.21	0	5	7	9	14
MC	1	Male	9.60	3.37	177	0.25	1	7	10	13	16
	1	Female	10.09	3.26	184	0.24	1	8	10	13	16
	3	Male	10.14	3.25	155	0.26	2	8	10	13	16
	3	Female	10.53	2.69	166	0.21	2	8	11	12	16
	5	Male	9.91	3.29	129	0.29	3	7	10	13	16
	5	Female	10.89	2.82	152	0.23	3	9	11	13	16
	7	Male	9.97	3.28	151	0.27	2	8	10	12	16
	7	Female	10.70	3.09	167	0.24	1	9	11	13	16

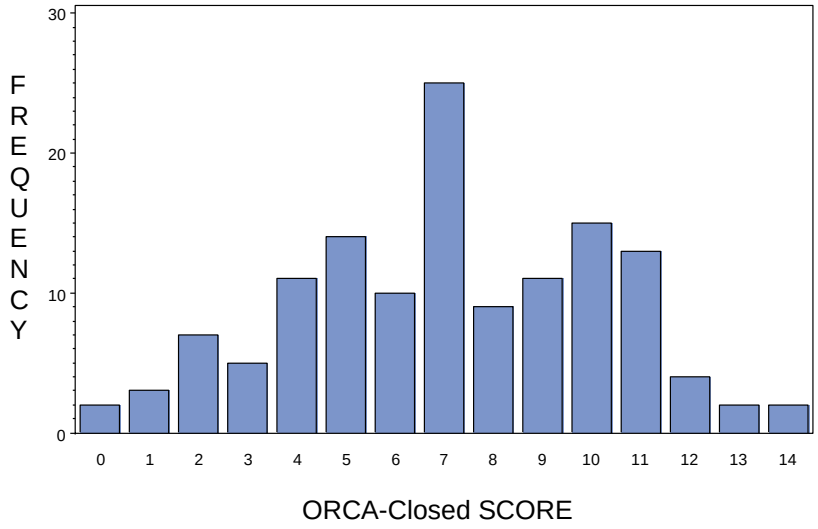
TABLE 3.3 Scores by Version separately by 1:1 Laptop/Non-Laptop

	Version	1:1 Laptop	Mean	Std Dev	Sample Size	Std Err	Min	Quartile 1	Median	Quartile 3	Max
ORCA-Closed	1	No	7.41	3.10	118	0.29	0	5	8	10	14
	1	Yes	7.15	2.74	153	0.22	1	5	7	9	14
	3	No	7.36	3.20	154	0.26	0	5	7	10	15
	3	Yes	6.81	3.12	147	0.26	1	5	7	9	14
	5	No	6.64	3.13	226	0.21	0	4	7	9	14
	5	Yes	6.20	3.26	142	0.27	0	4	6	8	14
	7	No	6.71	3.19	191	0.23	0	4	7	9	13
	7	Yes	6.31	2.97	150	0.24	1	4	6	9	14
MC	1	No	9.92	3.24	224	0.22	1	7	10	13	16
	1	Yes	9.72	3.44	137	0.29	1	7	10	13	16
	3	No	10.62	2.87	170	0.22	3	9	11	13	16
	3	Yes	10.03	3.06	151	0.25	2	8	11	12	16
	5	No	10.78	3.05	121	0.28	3	9	11	13	16
	5	Yes	10.18	3.08	160	0.24	3	8	10.5	13	16
	7	No	10.36	3.34	174	0.25	2	8	11	13	16
	7	Yes	10.35	3.03	144	0.25	1	9	11	12	16

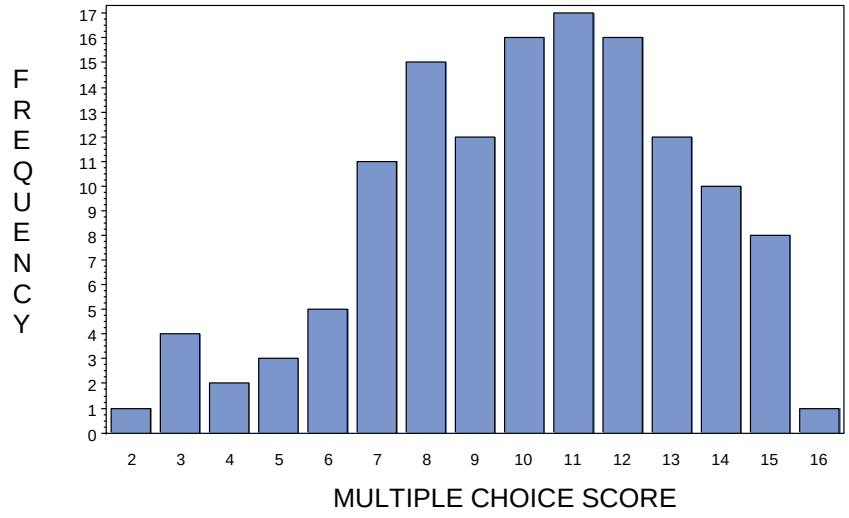
TABLE 3.4 Scores by Version Pairs

	Version Pair	Mean	Std Dev	Sample Size	Std Err	Min	Quartile 1	Median	Quartile 3	Max
TOTAL	1,3	17.22	5.46	133	0.47	3	14	17	21	28
	1,7	18.37	4.85	138	0.41	2	15	19	22	29
	3,1	16.91	6.04	161	0.48	2	13	17	21	29
	3,5	18.13	5.20	140	0.44	8	15	18	22	30
	5,3	17.40	5.26	188	0.38	4	14.5	17.5	21	29
	5,7	15.92	5.70	180	0.42	3	11	16.5	21	27
	7,1	16.51	5.68	200	0.40	3	12	17	21	28
	7,5	16.24	5.21	141	0.44	4	12	17	20	26
Difference	1,3	-2.98	2.93	133	0.25	-9	-5	-3	-1	5
	1,7	-3.57	2.62	138	0.22	-11	-5	-4	-2	3
	3,1	-2.63	3.00	161	0.24	-12	-5	-3	0	5
	3,5	-4.06	3.11	140	0.26	-11	-6	-4	-2	5
	5,3	-3.63	3.18	188	0.23	-12	-6	-3	-1.5	4
	5,7	-3.85	3.03	180	0.23	-11	-6	-4	-2	3
	7,1	-3.32	2.97	200	0.21	-11	-5	-3	-1	4
	7,5	-3.33	2.89	141	0.24	-10	-5	-3	-1	3

Topic Pair 1,3 - ORCA-Closed

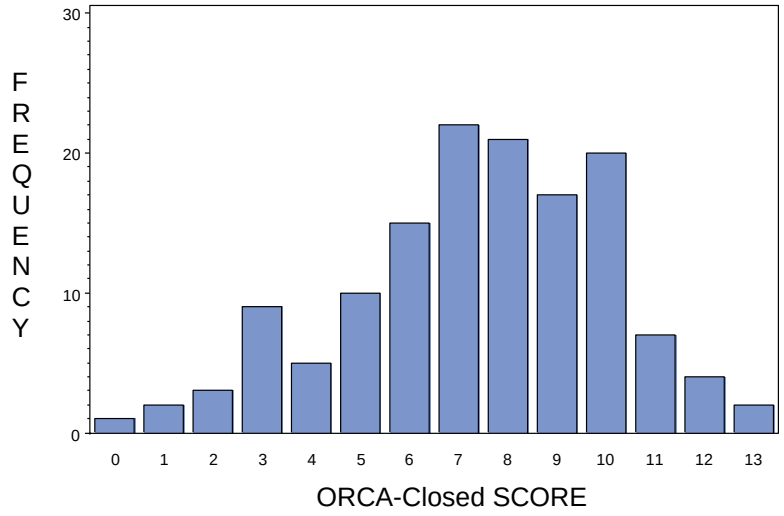


Topic Pair 1,3 - Multiple Choice

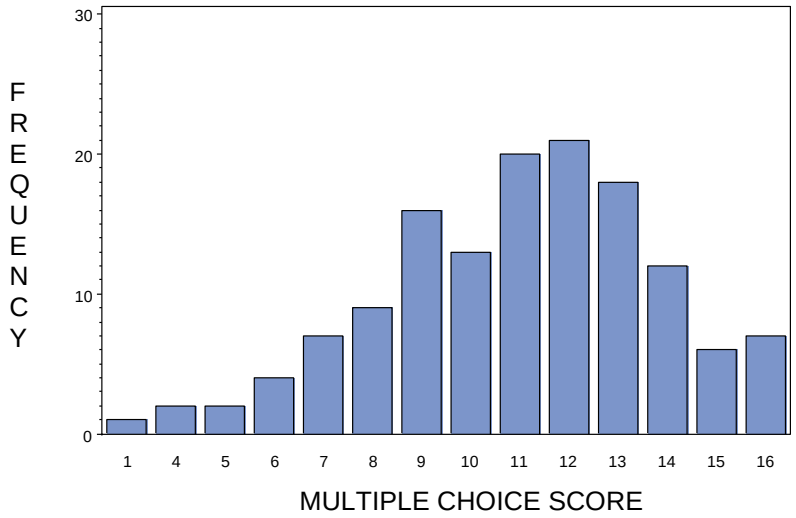


Topic Pair 1,3	Sample Size	Median Score	Mean	Std Dev	Std Err
Closed	133	7	7.12	3.13	0.27
MC	133	10	10.10	3.07	0.26

Topic Pair 1,7 - ORCA-Closed

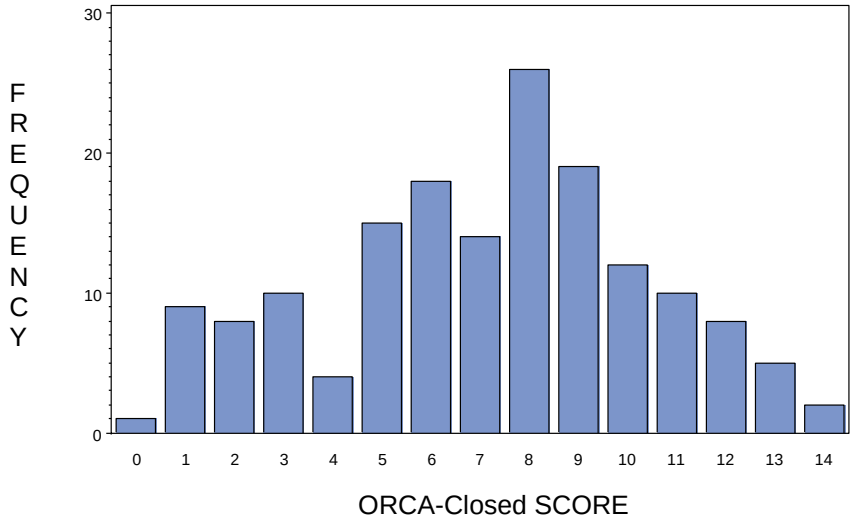


Topic Pair 1,7 - Multiple Choice

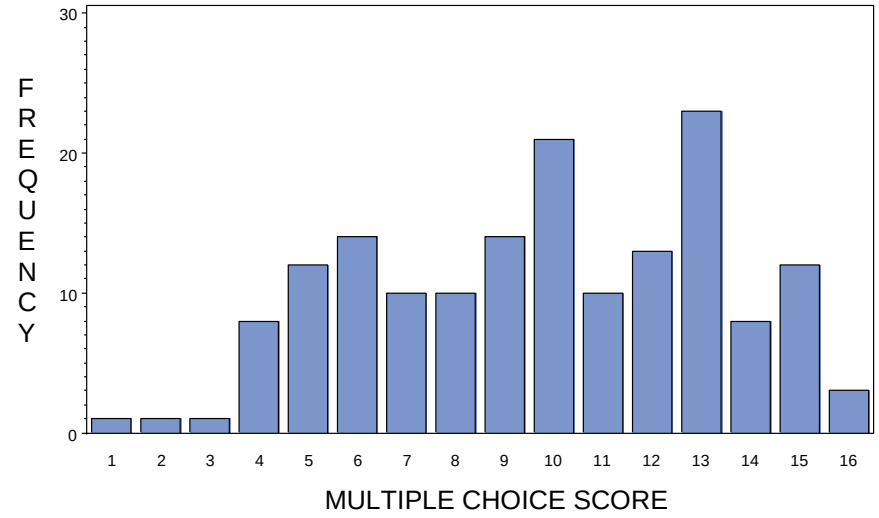


Topic Pair 1,7	Sample Size	Median Score	Mean	Std Dev	Std Err
Closed	138	8	7.40	2.66	0.23
MC	138	11	10.07	2.5	0.24

Topic Pair 3,1 - ORCA-Closed

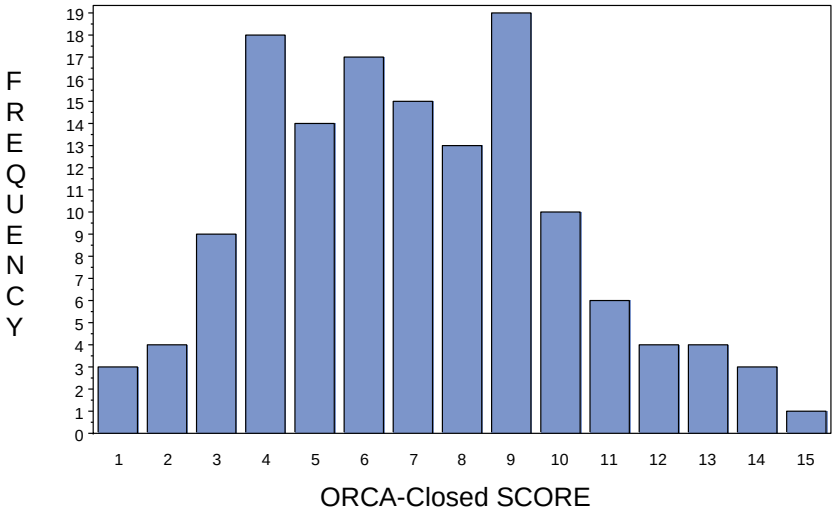


Topic Pair 3,1 - Multiple Choice

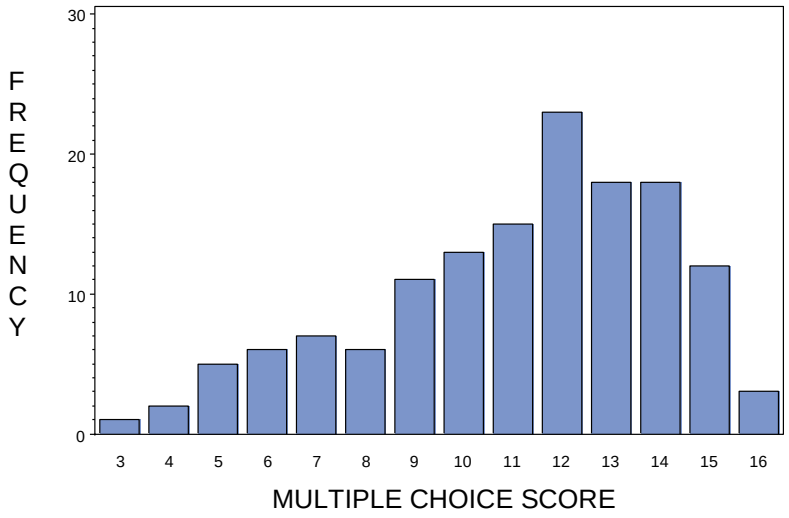


Topic Pair 3,1	Sample Size	Median Score	Mean	Std Dev	Std Err
Closed	161	8	7.14	3.25	0.26
MC	161	10	9.77	2.88	0.28

Topic Pair 3,5 - ORCA-Closed

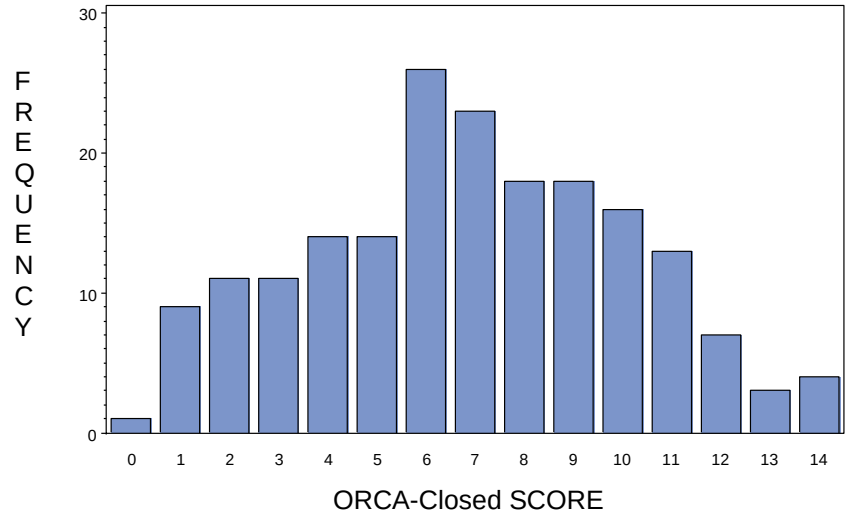


Topic Pair 3,5 - Multiple Choice

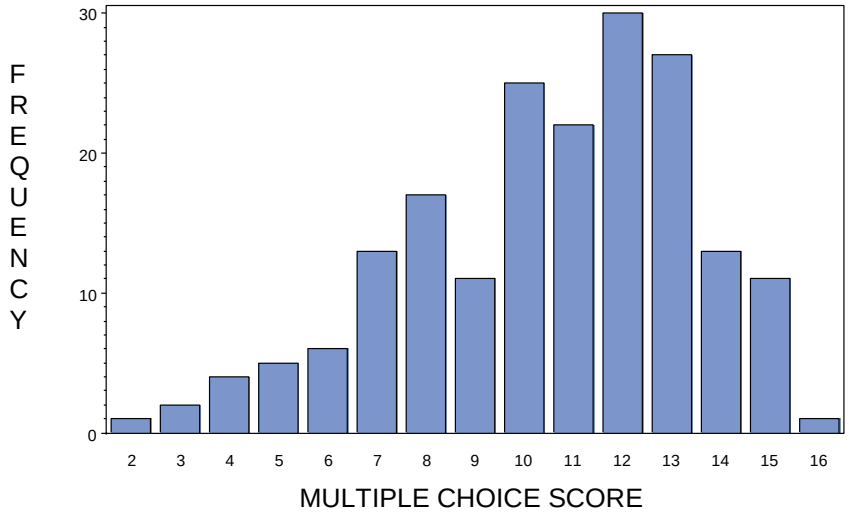


Topic Pair 3,5	Sample Size	Median Score	Mean	Std Dev	Std Err
Closed	140	7	7.04	3.08	0.26
MC	140	12	11.09	2.98	0.25

Topic Pair 5,3 - ORCA-Closed

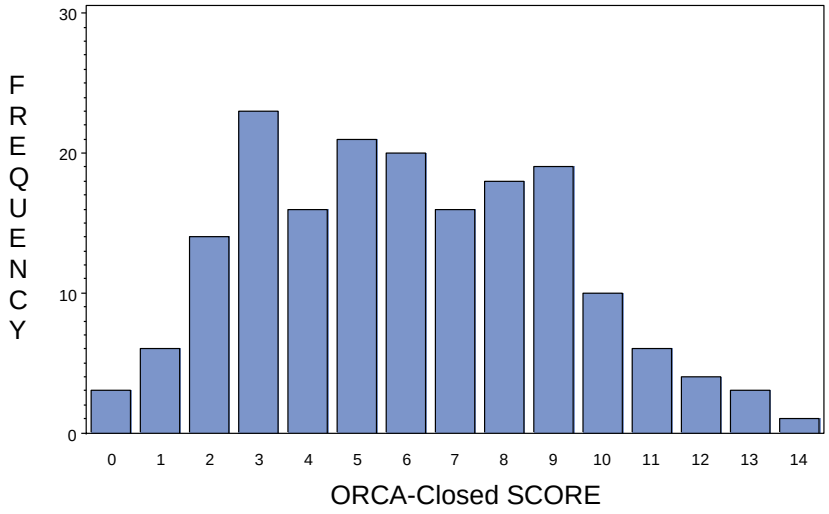


Topic Pair 5,3 - Multiple Choice

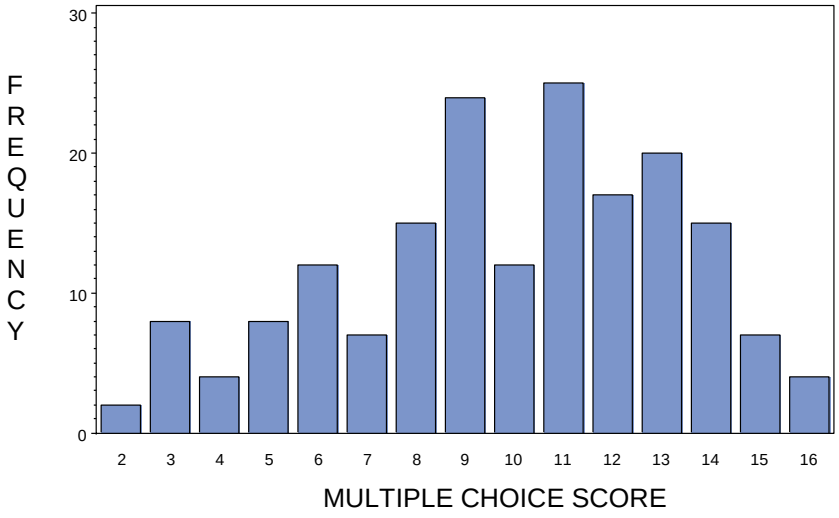


Topic Pair 5,3	Sample Size	Median Score	Mean	Std Dev	Std Err
Closed	188	7	6.89	3.23	0.24
MC	188	11	10.52	2.90	0.21

Topic Pair 5,7 - ORCA-Closed

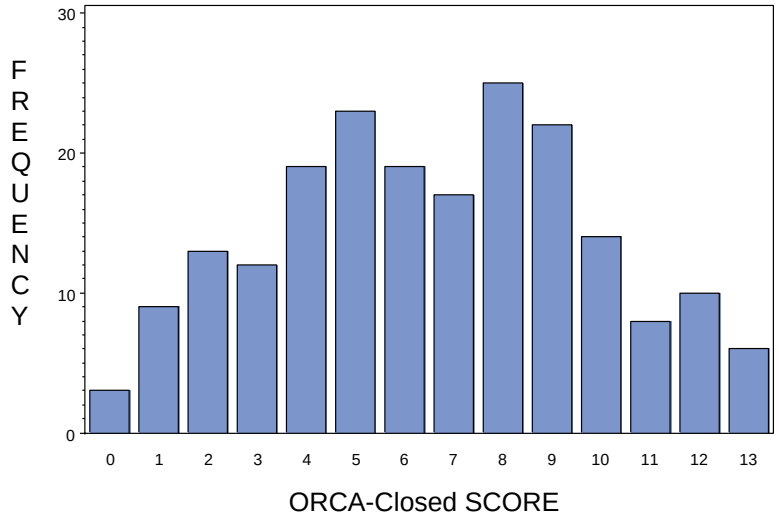


Topic Pair 5,7 - Multiple Choice

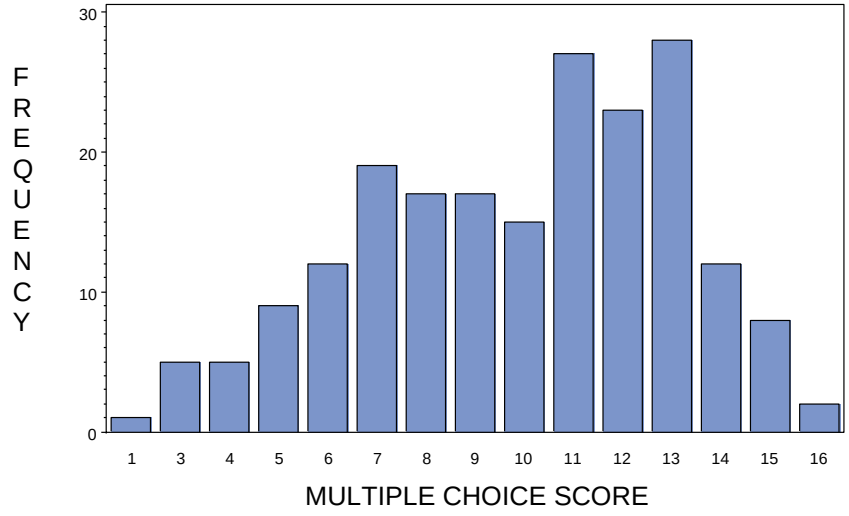


Topic Pair 5,7	Sample Size	Median Score	Mean	Std Dev	Std Err
Closed	180	6	6.03	3.08	0.23
MC	180	10	9.88	3.37	0.25

Topic Pair 7,1 - ORCA-Closed

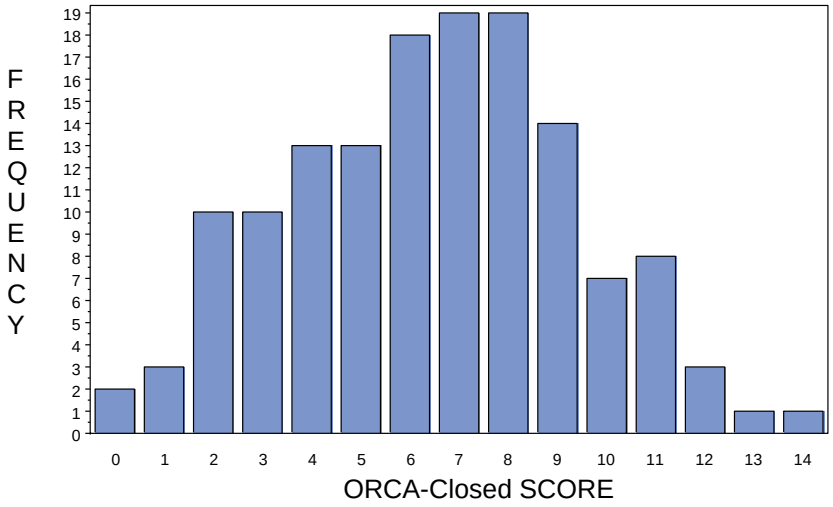


Topic Pair 7,1 - Multiple Choice

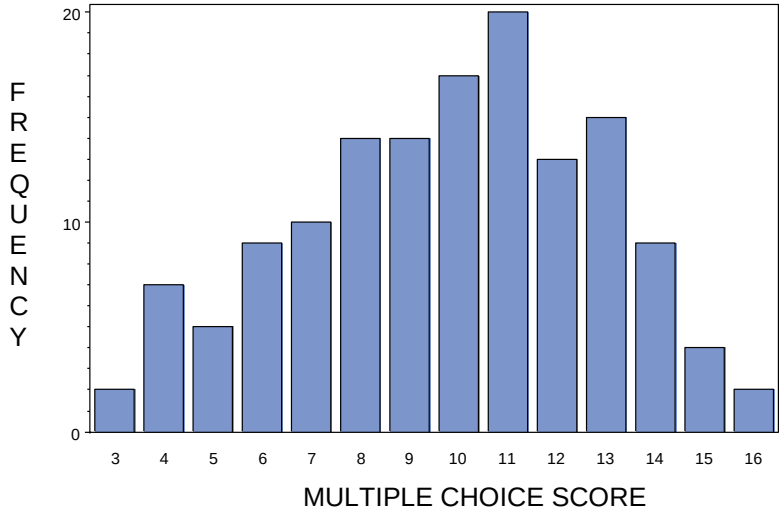


Topic Pair 7,1	Sample Size	Median Score	Mean	Std Dev	Std Err
Closed	200	7	6.60	3.22	0.23
MC	200	10	9.91	3.18	0.22

Topic Pair 7,5 - ORCA-Closed



Topic Pair 7,5 - Multiple Choice



Topic Pair 7,5	Sample Size	Median Score	Mean	Std Dev	Std Err
Closed	141	7	6.45	2.91	0.24
MC	141	10	9.79	3.04	0.26

4: *Question: What is the dimensionality of Multiple Choice? What is the dimensionality of ORCA?*

Assessing the dimensionality of an assessment directly addresses the question of the need for multiple distinct skills or latent traits required by the task and the efficiency of the assessment instrument in measuring these individually.

When a univariate model describes responses well, then either performance is driven by a single latent trait or else the instrument fails to elicit and separate the component skills or traits. (A third possibility is that the assessment is uniformly too easy or too hard to be able to distinguish among students by their performances on the test.) When a multidimensional model is needed to adequately describe performance, then multiple traits are indicated.

As a measure of online reading or research ability, the assessments constructed as part of the ORCA project were designed to elicit separate measurements of the four components and to determine the extent to which the LESC components relate to separate skills. Therefore by fitting both unidimensional item response (UIRT) and multidimensional item response (MIRT) models and comparing the adequacy of these models, the nature of the task and the assessment can be evaluated.

The first big study as part of the ORCA project, ORCA I, was a cross-sectional study of three formats for the assessment: ORCA-Open (with access to the open internet), the otherwise identical ORCA-Closed (with access only to an extensive synthetic internet), and the on-screen Multiple Choice version of the same tasks. Each student took two assessments in a single format; and comparative analysis of performance by format was made by comparing groups of students. By contrast, in ORCA II, the subject of this technical report, each student took one ORCA-Closed assessment and one Multiple Choice assessment. Thus comparative analysis of performance by format is conducted within student.

Results from ORCA I indicated that ORCA-Open and ORCA-Closed were indeed multidimensional assessments and the dimensions accorded quite well with the presumed components, LESC. Analysis also showed that Multiple Choice, if not truly a unidimensional assessment, was very well approximated by a unidimensional model and any higher dimension drivers were very subtle, if present.

In between ORCA I and ORCA II minor changes were made to the ORCA-Closed assessment, largely to clarify scoring. However, several items in the Multiple Choice assessment were revised to increase the level of difficulty either by exchanging texts referred to in the items or by including better distractors. As a consequence, the psychometric properties of the Multiple Choice assessment were improved. At the same time as is noted below the Multiple Choice assessment for ORCA II became unambiguously unidimensional.

The alternative models evaluated for both ORCA I and ORCA II are shown in Table 4.1 for each dimensionality up to four. Both a 4-dimensional item response model (4-dimensional MIRT) and a Bifactor model with 4 subdimensions were fitted to the ORCA data because, although they both fit the same dimensionality (and have the same overall results and interpretations) the Bifactor model fits orthogonal subdimensions whereas the 4-dimensional MIRT fits the dimensions

specified by the proposed components, in this case LESC. Therefore the two kinds of model allow different insight sometimes leading to clearer inferences with one than another.

As Tables 4.2 and 4.3 show for ORCA I and ORCA II respectively, ORCA-Closed is clearly multidimensional, essentially 4-dimensional, according to all of the goodness-of-fit criteria. By contrast for both ORCA I and ORCA II, the Multiple Choice assessment is unidimensional; for each goodness-of-fit criterion differences are trivial among the values for all dimensions from one to four. This is further borne out by the high correlations among component subscores on the Multiple Choice assessment and the high correlations among dimensions for Multiple Choice in Tables 4.6 and 4.7. Note that for Multiple Choice the higher correlations in Table 4.7 (ORCA II) than in Table 4.6 (ORCA I) reflect changes made to this format prior to administration of ORCA II. The pattern of only modest correlations for ORCA-Closed is similar between ORCA I and ORCA II.

Tables 4.4 and 4.5 are included for completeness as these provide the information side by side for the Bifactor model with 4 subdimensions and the 4-dimensional MIRT. Inferences about the dimensionality of the ORCA-Closed and the Multiple Choice are the same for both types of 4-dimensional model.

From Tables 4.6, 4.7 and 4.8, the concordance of the dimensions of ORCA-Closed is seen to be primarily aligned with the components LESC, although it is less than perfect. For ORCA II the General factor of the Bifactor Model with 4 subdimensions still contributes to the model with a variance that is slightly greater than that of a unidimensional model (UIRT) and approximately the same for the General factor of the Multiple Choice assessment.

Table 4.1 Alternative Models proposed for ORCA Project Assessments

Model	Dimension of Model	Model Structure
4-Dimension MIRT	L|E|S|C	First 4 items (Locate) load on 1st latent dimension Next 4 items (Synthesize) load on 2nd latent dimension Third 4 items (Evaluate) load on 3rd latent dimension. Last 4 items (Communicate) load on 4th latent dimension. The four latent dimensions are correlated.
UIRT	LESC	All 16 items load on one single latent dimension.
2-Dimension MIRT	LES|C	First 12 items (Locate, Evaluate & Synthesize) load on 1st latent dimension Last 4 items (Communicate) load on 2nd latent dimension. The two latent dimensions are correlated.
3-Dimension MIRT	L|ES|C	First 4 items (Locate) load on 1st latent dimension Next 8 items (Synthesize & Evaluate) load on 2nd latent dimension Last 4 items (Communicate) load on 3rd latent dimension. The three latent dimensions are correlated.

Bifactor Model 4 Subdimensions	G L|E|S|C	All 16 items load on the general dimension. In addition, First 4 items (Locate) load on 1st latent subdimension Next 4 items (Synthesize) load on 2nd latent subdimension Third 4 items (Evaluate) load on 3rd latent subdimension. Last 4 items (Communicate) load on 4th latent subdimension. The latent dimensions are all orthogonal (i.e., uncorrelated).

Table 4.2 Goodness-of-Fit Statistics for Alternative Models for ORCA I by Format

Format	Model	-2Loglikelihood	Number of Free Parameters	AIC	BIC	Δ-2LL	DF
ORCA-Open	4-dimension MIRT	5165	38	5241	5393	--	--
	3-dimension MIRT	5187	35	5257	5397		3
	2-dimension MIRT	5353	33	5419	5551		5
	UIRT	5402	32	5466	5594	237	6
ORCA-Closed	4-dimension MIRT	5988	38	6064	6216	--	--
	3-dimension MIRT	5995	35	6065	6205		3
	2-dimension MIRT	6146	33	6212	6344		5
	UIRT	6192	32	6256	6384	205	6
ORCA-MC	4-dimension MIRT	6955	38	7031	7183	--	--
	3-dimension MIRT	6967	35	7037	7177		3
	2-dimension MIRT	6988	33	7054	7186		5
	UIRT	6998	32	7062	7191	43.52	6

Table 4.3 Goodness-of-Fit Statistics for Alternative Models for ORCA II by Format

Format	Model	-2Loglikelihood	Number of Free Parameters	AIC	BIC	Δ-2LL	DF
ORCA-Closed	4-dimension MIRT	13130	38	13206	13358	--	--
	3-dimension MIRT	13200	35	13270	13410	70	3
	2-dimension MIRT	13325	33	13391	13523	195	5
	UIRT	13434	32	13498	13626	304	6
ORCA-MC	4-dimension MIRT	13983	38	14059	14211	--	--
	3-dimension MIRT	13989	35	14059	14199	16	3
	2-dimension MIRT	13987	33	14053	14185	4	5
	UIRT	13982	32	14046	14174	-1	6

Table 4.4 ORCA I Goodness-of-Fit Statistics for 4-Dimension MIRT and Bifactor Models by Format

Format	Model	-2Loglikelihood	Number of Free Parameters	AIC	BIC	Δ (-2LL)	DF
ORCA-Open	4-dimension MIRT	5165	38	5241	5393	--	--
	Bifactor Model 4 Subdimensions	5142	48	5238	5430	23	10
ORCA-Closed	4-dimension MIRT	5988	38	6064	6216	--	--
	Bifactor Model 4 Subdimensions	5972	48	6068	6260	16	10
MC	4-dimension MIRT	6955	38	7031	7183	--	--
	Bifactor Model 4 Subdimensions	6942	48	7038	7230	13	10

Table 4.5 ORCA II Goodness-of-Fit Statistics for 4-Dimension MIRT and Bifactor Models by Format

Format	Model	-2Loglikelihood	Number of Free Parameters	AIC	BIC	Δ (-2LL)	DF
ORCA-Closed	4-dimension MIRT	13130	38	13206	13358	--	--
	Bifactor Model 4 Subdimensions	13044	48	13140	13332	86	10
MC	4-dimension MIRT	13983	38	14059	14211	--	--
	Bifactor Model 4 Subdimensions	13950	48	14046	14238	33	10

Table 4.6 ORCA I Correlation Matrix of Dimensions in 4-dimension MIRT Model

ORCA-Open					ORCA-Closed					ORCA-MC				
	L	E	S	C		L	E	S	C		L	E	S	C
L	**1**				L	**1**				L	**1**			
E	0.22	**1**			E	0.51	**1**			E	0.39	**1**		
S	0.36	0.61	**1**		S	0.43	0.78	**1**		S	0.78	0.75	**1**	
C	0.46	0.39	0.50	**1**	C	0.50	0.60	0.51	**1**	C	0.60	0.67	0.76	**1**

Note: L=Locate E=Evaluate S=Synthesize C=Communicate.
Correlations were computed as direct estimates of latent scales using IRTPRO2.1.

Table 4.7 ORCA II Correlation Matrix of Dimensions in 4-dimension MIRT Model

ORCA-Closed					ORCA-MC				
	L	E	S	C		L	E	S	C
L	**1**				L	**1**			
E	0.41	**1**			E	0.81	**1**		
S	0.51	0.58	**1**		S	0.83	0.85	**1**	
C	0.34	0.44	0.37	**1**	C	0.85	0.89	0.80	**1**

Table 4.8 ORCA I Latent Dimensions Variance Estimates for Alternative Models

Model	Dimension	ORCA-Open	ORCA-Closed	ORCA-MC
UIRT	UIRT	0.67	1.36	0.75
4-Dimension MIRT	MIRT-L	0.92	2.12	0.70
	MIRT-E	0.65	0.85	0.52
	MIRT-S	0.64	1.50	0.66
	MIRT-C	0.55	1.13	1.17
Bifactor Model 4 subdimensions	BiFactor-G	0.66	0.80	0.74
	Bifactor-SubL	1.21	1.34	0.33
	Bifactor-SubE	0.82	0.56	0.41
	Bifactor-SubS	0.54	0.77	0.44
	Bifactor-SubC	0.59	0.61	0.68

Note: Variances in this table were computed as direct estimates of latent scales using IRTPRO2.1.

Table 4.9 ORCA II Latent Dimensions Variance Estimates for Alternative Models

Model	Dimension	ORCA-Closed	ORCA-MC
UIRT	UIRT	0.47	0.51
4-Dimension MIRT	MIRT-L	0.52	0.62
	MIRT-E	1.12	1.58
	MIRT-S	1.11	1.5
	MIRT-C	1.63	1.18
Bifactor Model 4 subdimensions	BiFactor-G	0.59	0.58
	Bifactor-SubL	0.52	1.26
	Bifactor-SubE	0.69	0.68
	Bifactor-SubS	1.08	0.54
	Bifactor-SubC	1.3	0.67

5: *Question: How do students perform differently on Multiple Choice and ORCA? Are there any specific differences for specific components?*

Relationship between individual students' scores on ORCA- Closed and on MC
As seen in the Table 5.1, there is a wide range of the bivariate scores (ORCA-Closed score and Multiple Choice score). From this table the middle group of students can be identified, i.e., those whose scores on each assessment fell within the (approximate) middle half of scores for that format. The shaded boxes represent identical scores on the two formats. A total of 99 students actually had the same score on the ORCA-Closed as on the Multiple Choice assessment. The middle group included students who scored between 4 and 10 on the ORCA, and between 8 and 13 for Multiple Choice; of these, most scored higher on Multiple Choice than on ORCA-Closed. Overall, only 120 of the1281 students scored higher on ORCA-Closed than on Multiple Choice. In looking at the three areas of Table 5.1, (the low-scoring group, the middle group, and the higher-scoring group) patterns of score differentials are quite distinct. For the low-scoring group, the differentials were too small to discern – partially due to the few scorers and also due to the small score range possible. The middle-range and the higher-scoring groups behaved quite differently. In terms of Multiple Choice, the middle-range group's scores were 2.5-3 points lower for ORCA-Closed scores; and approximately 4 points lower for the higher-scoring group. What these data indicate is that, while on the whole students did better on the Multiple Choice test than the ORCA, the disparity between scores as Multiple Choice scores increased as scores were overall higher.

Table 5.1 Students' Score Pairs for ORCA-Closed and Multiple Choice

Multiple Choice Score

ORCA - Closed Score	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
0				2	2		3	1	1	4							
1		2	1	1	6	4	5	4	9	2	4	2	3	1			
2		1	2	9	4	7	8	6	6	8	5	8	2	3	1		
3				4	6	7	8	13	14	11	10	6	1	6	3		
4				2	4	9	12	13	6	8	5	15	13	12	1		
5			2	3	6	8	8	10	10	15	11	22	15	8	5	1	
6				1	3	4	6	11	15	21	20	23	14	12	9	4	
7				1	2	3	10	9	20	16	20	16	28	12	6	7	1
8						4	5	6	11	13	17	18	23	23	13	13	3
9					1	2	1	5	4	9	18	17	18	36	15	9	4
10						1	1	1	6	9	12	12	18	18	14	10	2
11							1	3	1	3	6	9	7	11	17	7	6
12								2			1	6	10	6	7	7	5
13											2	1	3	9	3	6	2
14											1	1	1	4	3	3	
15																1	
16																	

Patterns of score differences on ORCA-Closed and on MC, by Component
Score comparisons between ORCA and Multiple Choice within student are shown for each component subscore in Table 5.2. Differences in the spread of scores groups differed for the four LESC components. To consider whether ORCA or Multiple Choice scoring was higher, lower, or equal, we examined those scores that differed by more than one point from equality for the two formats, with Multiple Choice higher than ORCA-Closed.

For *Locate*, it was found that students scored better on ORCA 286 times out of the 1281 students, or 22.3%. ORCA and Multiple Choice scored equally 325 times (25.4%); and Multiple Choice as a higher score occurred 670 times (52.3%). Looking deeper at increments greater than a single point, Multiple Choice scores more than one point above equal values were seen 356 times, or 27.8% of the time.

For *Evaluate*, students scored better on ORCA than Multiple Choice only 108 times (8.4%); and ORCA and Multiple Choice were scored equally well 266 times (20.8%). Multiple Choice scores were higher 907 times (70.8%), with 543 students scoring more than one point above equivalence (42.4%).

For *Synthesize*, students fared better on ORCA 407 times (31.8%); and fared equally well on both 376 times (29.4). Of the remaining 498 occurrences where students scored higher on Multiple Choice (38.9% of the time), 242 (18.9% of the total) were scores at least one point above the ORCA-Multiple Choice equivalence line.

For *Communicate*, students on scored higher on ORCA only 96 times (7.5%); and only 180 times (14.1%) for equal scoring. Students thus scored higher on Multiple Choice 1005 times (78.5%), with 636 (49.7%) having a score disparity of more than one point.

The first conclusion to be drawn is that on each of the component subscores the preponderance of students scored higher on Multiple Choice than ORCA. For each of the component subscores, approximately half the students had a Multiple Choice score that was equal to or one point higher than the ORCA-Closed score. Patterns for the other half of the students differed greatly for the different components. The biggest contrast to the overall trend was for Synthesize with the highest proportion of students (31.8%) performing better on ORCA-Closed than on Multiple Choice. At the other extreme, few (even very few: 8.4% for Evaluate and 7.5% for Communicate) students scored higher on ORCA-Closed for the other three LESC components. The proportion of students with identical score on ORCA-Closed and on Multiple Choice was fairly comparable across all LESC components, generally including 20-30% of students.

The most dramatic differences between Multiple Choice and ORCA-Closed scores occurred with Evaluate and Communicate. In both cases, the proportion of students scoring higher on ORCA than Multiple Choice was less than 10%; and both had a proportion of students scoring higher on Multiple Choice than ORCA of at least 70%, indicating a likelihood of a pervasive weakness in students' online abilities in these two areas.

Unlike Evaluate and Communicate, which depicted extremes between students' ORCA and Multiple Choice scores, students' scores distributions for Synthesize were consistent with a

balanced distribution of abilities, as there was only a 9.5% difference between the highest proportion (Multiple Choice as higher) and the lowest (Multiple Choice equivalent to ORCA). Also, this was the only component with more than a quarter of students scoring higher for ORCA than for Multiple Choice. Two possible reasons for this could be that there may be greater similarity in the tasks between the two formats or that the cognitive skills involved in Synthesize are less of a function of online or offline reading skills than for the other components.

Finally, of the students who scored higher on Multiple Choice, there was a relatively consistent proportion that scored more than one point higher. For Locate the percentage was 53.1%; for Evaluate it was 59.9%; for Synthesize 48.6%; and for Communicate 63.3%. This would seem to be a general descriptor of the student population indicating that among students who perform better on a traditional task-based test, approximately half do significantly less well on a performance-based assessment like ORCA-Closed.

TABLE 5.2 Students' Component Score Pairs for ORCA-Closed and Multiple Choice Scores

Locate

ORCA - Closed Score \ Multiple Choice Score	0	1	2	3	4
0	20	50	70	62	20
1	19	61	84	83	50
2	15	55	107	119	71
3	4	25	72	99	61
4	3	13	35	45	38
Mean	1.20	1.46	1.78	1.96	2.20

Synthesize

ORCA - Closed Score \ Multiple Choice Score	0	1	2	3	4
0	11	40	47	24	17
1	13	47	67	70	32
2	14	37	67	67	52
3	13	33	64	110	82
4	6	25	73	129	141
Mean	1.82	1.76	2.15	2.63	2.92

Evaluate

ORCA - Closed Score \ Multiple Choice Score	0	1	2	3	4
0	15	38	36	46	20
1	30	83	132	186	129
2	9	27	90	116	126
3	1	4	19	48	76
4		2	4	12	30
Mean	0.93	1.02	1.37	1.50	1.91

Communicate

ORCA - Closed Score \ Multiple Choice Score	0	1	2	3	4
0	40	118	169	131	70
1	7	43	119	101	78
2	6	28	56	83	87
3	1	9	32	37	49
4		1	4	8	4
Mean	0.41	0.65	0.90	1.14	1.44

6: *Question: How do Component Scores differ between ORCA-Closed and Multiple Choice with respect to overall influential factors and with respect to version? What relationships are evident among component scores?*

To make comparisons for each LESC component separately with respect to gender, laptop condition, and test version, average scores and differences between ORCA formats were calculated. Average scores for each component for each version are shown in Table 6.1.

Gender contributed significantly to scores for three of the four LESC components of ORCA-Closed scores and for a different set of three components for Multiple Choice scores. On ORCA-Closed, males scored significantly lower than females for Locate, Evaluate and Synthesize; but there was only a large magnitude of effect for Synthesize. In regard to Communicate, the significance of the difference was only marginal. For Multiple Choice, score difference between males and females for Evaluate, Synthesize, and Communicate were significant; in this format, Locate was only marginally significant. In all cases for all components, males scored lower than did females. Lastly, examining the (within student) Differences for Synthesize showed that ORCA-Closed was more difficult for males, while there was little difference in scoring for females between the two formats. No patterns related to gender were evident for within-student differences for Locate, Evaluate, and Communicate.

The laptop condition (1:1 laptops in classrooms or not) was analyzed in the same fashion as was gender. For the ORCA-Closed, non-laptop students scored significantly higher than did laptop students for the Evaluate and Synthesize components. For Communicate the difference due to laptop condition was non-significant, as was true for Locate. But while the latter was a non-significant difference, it did display laptop students scoring higher than non-laptop students, a reversal of the trend for other components. For Multiple Choice, there were no truly significant differences between non-laptop students and laptop students for any of the components. However for Evaluate and Synthesis results were marginally significant with non-laptop students scoring higher than laptop students. Comparing test difficulty for each of the components gives no evidence of difference between laptop and non-laptop conditions.

Next the importance of the version of the test was considered. For ORCA-Closed, average scores for all versions for Locate ranged significantly: from 1.58 to 2.2. Scores for Synthesize also varied significantly, ranging from 2.15 to 2.7. For both Evaluate and Communicate, score ranges with a single exception were narrow. Average scores did not differ significantly, with average scores being within 0.1 of each other. The exception was a singular distinct, low average value (1.39) for version 7 for Evaluate. Focusing on the Multiple Choice format, there were generally significant score ranges. For Locate, the range was from 2.31 to 2.60, a range of .29. For Synthesize, there was a .41 range (2.37 to 2.78). Communicate and Evaluate were both clustered, lacking significant range disparities; but for Evaluate the average score was 2.57 for version 1 - again distinct and lower than for the other versions.

Table 6.1 LESC Component Scores by Version

		Version	Mean	Std Dev	Sample Size	Std Err	Min	Quartile 1	Median	Quartile 3	Max
Closed	LESC	1	7.26	2.90	271	0.18	0	5	7	10	14
		3	7.09	3.17	301	0.18	0	5	7	9	15
		5	6.47	3.18	368	0.17	0	4	6	9	14
		7	6.54	3.10	341	0.17	0	4	7	9	14
	Locate	1	1.93	1.19	271	0.07	0	1	2	3	4
		3	2.20	1.26	301	0.07	0	1	2	3	4
		5	1.71	1.16	368	0.06	0	1	2	2	4
		7	1.58	1.25	341	0.07	0	0	2	3	4
	Evaluate	1	1.60	0.99	271	0.06	0	1	1	2	4
		3	1.55	1.04	301	0.06	0	1	1	2	4
		5	1.53	0.97	368	0.05	0	1	1	2	4
		7	1.39	0.91	341	0.05	0	1	1	2	4
	Synthesis	1	2.70	1.32	271	0.08	0	2	3	4	4
		3	2.40	1.41	301	0.08	0	1	3	4	4
		5	2.15	1.31	368	0.07	0	1	2	3	4
		7	2.52	1.34	341	0.07	0	1	3	4	4
	Communicate	1	1.03	1.04	271	0.06	0	0	1	2	4
		3	0.95	1.07	301	0.06	0	0	1	2	4
		5	1.08	1.09	368	0.06	0	0	1	2	4
		7	1.05	1.05	341	0.06	0	0	1	2	4
MC	LESC	1	9.85	3.32	361	0.17	1	7	10	13	16
		3	10.34	2.98	321	0.17	2	8	11	13	16
		5	10.44	3.08	281	0.18	3	8	11	13	16
		7	10.36	3.20	318	0.18	1	8	11	13	16
	Locate	1	2.46	1.13	361	0.06	0	2	3	3	4
		3	2.60	0.99	321	0.06	0	2	3	3	4
		5	2.38	1.09	281	0.07	0	2	2	3	4
		7	2.31	1.18	318	0.07	0	1	2	3	4
	Evaluate	1	2.57	1.18	361	0.06	0	2	3	4	4
		3	2.72	1.12	321	0.06	0	2	3	4	4
		5	2.81	1.13	281	0.07	0	2	3	4	4
		7	2.78	1.11	318	0.06	0	2	3	4	4
	Synthesis	1	2.37	1.16	361	0.06	0	1	2	3	4
		3	2.50	1.13	321	0.06	0	2	3	3	4
		5	2.74	1.10	281	0.07	0	2	3	4	4
		7	2.78	1.12	318	0.06	0	2	3	4	4
	Communicate	1	2.45	1.08	361	0.06	0	2	2	3	4
		3	2.52	1.12	321	0.06	0	2	3	3	4
		5	2.51	1.15	281	0.07	0	2	3	4	4
		7	2.49	1.16	318	0.06	0	2	3	3	4

TABLE 6.2 LESC Component Scores by Version Pairs

		Version Pair	Mean	Std Dev	Sample Size	Std Err	Min	Quartile 1	Median	Quartile 3	Max
TOTAL		1,3	17.22	5.46	133	0.47	3	14	17	21	28
		1,7	18.37	4.85	138	0.41	2	15	19	22	29
		3,1	16.91	6.04	161	0.48	2	13	17	21	29
		3,5	18.13	5.20	140	0.44	8	15	18	22	30
		5,3	17.40	5.26	188	0.38	4	14.5	17.5	21	29
		5,7	15.92	5.70	180	0.42	3	11	16.5	21	27
		7,1	16.51	5.68	200	0.40	3	12	17	21	28
		7,5	16.24	5.21	141	0.44	4	12	17	20	26
Difference	Score	1,3	-2.98	2.93	133	0.25	-9	-5	-3	-1	5
		1,7	-3.57	2.62	138	0.22	-11	-5	-4	-2	3
		3,1	-2.63	3.00	161	0.24	-12	-5	-3	0	5
		3,5	-4.06	3.11	140	0.26	-11	-6	-4	-2	5
		5,3	-3.63	3.18	188	0.23	-12	-6	-3	-1.5	4
		5,7	-3.85	3.03	180	0.23	-11	-6	-4	-2	3
		7,1	-3.32	2.97	200	0.21	-11	-5	-3	-1	4
		7,5	-3.33	2.89	141	0.24	-10	-5	-3	-1	3
	Locate	1,3	-0.63	1.29	133	0.11	-3	-2	-1	0	3
		1,7	-0.84	1.57	138	0.13	-4	-2	-1	0	4
		3,1	-0.27	1.42	161	0.11	-4	-1	0	1	3
		3,5	-0.61	1.65	140	0.14	-4	-2	-1	0.5	4
		5,3	-0.52	1.45	188	0.11	-4	-1.5	0	0	3
		5,7	-0.66	1.37	180	0.10	-4	-2	-1	0	3
		7,1	-0.84	1.53	200	0.11	-4	-2	-1	0	4
		7,5	-0.43	1.36	141	0.11	-3	-1	0	1	2
	Evaluate	1,3	-1.07	1.33	133	0.12	-4	-2	-1	0	2
		1,7	-1.31	1.16	138	0.10	-4	-2	-1	-1	2
		3,1	-1.10	1.36	161	0.11	-4	-2	-1	0	3
		3,5	-1.34	1.35	140	0.11	-4	-2	-1	0	2
		5,3	-1.13	1.14	188	0.08	-4	-2	-1	0	1
		5,7	-1.25	1.25	180	0.09	-4	-2	-1	0	2
		7,1	-1.15	1.18	200	0.08	-4	-2	-1	0	2
		7,5	-1.27	1.25	141	0.11	-4	-2	-1	-1	3
	Synthesis	1,3	0.19	1.54	133	0.13	-4	-1	0	1	4
		1,7	0.28	1.26	138	0.11	-3	0	0	1	3
		3,1	0.38	1.50	161	0.12	-4	-1	0	2	3
		3,5	-0.56	1.45	140	0.12	-4	-2	0	0	4
		5,3	-0.60	1.53	188	0.11	-4	-2	-1	0	4
		5,7	-0.64	1.36	180	0.10	-4	-2	-1	0	3
		7,1	0.11	1.41	200	0.10	-4	-1	0	1	3
		7,5	-0.35	1.41	141	0.12	-4	-1	0	0	3

	Communicate	1,3	-1.47	1.38	133	0.12	-4	-2	-1	-1	3
		1,7	-1.70	1.26	138	0.11	-4	-3	-2	-1	1
		3,1	-1.63	1.32	161	0.10	-4	-2	-2	-1	2
		3,5	-1.54	1.42	140	0.12	-4	-3	-2	-1	3
		5,3	-1.39	1.38	188	0.10	-4	-2	-1	-1	2
		5,7	-1.30	1.27	180	0.09	-4	-2	-1	0	2
		7,1	-1.44	1.19	200	0.08	-4	-2	-1	-1	2
		7,5	-1.28	1.33	141	0.11	-4	-2	-1	0	2

7: *Question: What factors drive students' performance (with Multiple Choice and ORCA separately) from among test attributes, school attributes, student attributes, and student computer familiarity?*

Factors that could influence scores were analyzed for each format in four ways: Multiple Choice, ORCA-Closed, TOTAL, and Differences between Multiple Choice and Closed scores for each student. The first three give information about factors affecting performance; the score differences investigates factors associated with preference or skill differences between the two formats. Candidate factors are given for each model fitted to the data for 1281 students. Factors of specific interest include: 1:1 laptop in schools (state 1) or no laptop (state 2), gender, version, socioeconomic status (%FRPL eligible), order of test, offline reading measurement (ORM), prior knowledge, reported hours of computer use per day personal computer at home, number of computers in the home, reported hours of computer use per day and self-assessed expertise of computer use.

For Multiple Choice, the list of potential factors contained both significant and non-significant factors. For gender, there was a significant difference between female and male scoring ($p < .0001$) with females scoring on average approximately 0.6 points higher than men. Version of the test was also significant ($p = .0236$). The average scores for different versions ranged from 9.85 to 10.43 points, a .58-point variation. The most important factor was the score on the Offline Reading Measure. ORM was a very highly significant factor ($p < .0001$) and accounted for 686.7 out of the total 1099.9 (62.4%) of the total Type III SS for the model that incorporated all potential factors listed above. The impact of Prior Knowledge impact was also significant ($p < .0001$), accounting for 188 of the 1099.4 (17.1%) Type III SS. Next most important was self-assessed Expertise with computers, which accounted for 120.5 (10.9%). Other potential factors were non-significant: FRPL, having use of a laptop or not, the use of computers at home, the number of computers at home, and hours spent daily on the computer. Neither the day of the test nor the order of MC and ORCA-Closed was a significant factor in this model. Thus no learning effect was evident on Multiple Choice for students taking the assessment in two different formats.

For ORCA-Closed test, more factors were important. Still ORM was the single most important factor, again very highly significant ($p < .0001$), accounting for 499.3 of the 966.2 (51.7%) Type III SS. Gender appeared to be even more important for ORCA-Closed than for Multiple Choice, again highly significant ($p < .0001$) but with females on average scoring 1.2 points higher than males. The version of the test also mattered ($p < .0001$), with scores ranging from 6.48 to 7.26 (0.75 point range). This difference attributed to gender is explored later in some detail. Of lesser importance although still statistically significant, %FRPL-eligible ($p = .0235$) accounted for 15.9 (1.6%); and prior knowledge ($p = .0009$) accounted for 34.4 (3.1%) of the Type III SS. As with Multiple Choice, the order of the test was non-significant., again indicating that there was no learning effect on the ORCA-Closed performance when two assessments were in different formats.

In the internet setting of ORCA-Closed, the laptop condition was also significant ($p = .0044$), with non-laptop students faring .33 points better than laptop students. Hours spent on the computer at home ($p = .0105$) accounted for 81 of the 966.2 (8.4%); and self-reported expertise with a computer ($p < .0001$) accounted for 70.9 (7.3%). The number of computers in the student's home was not a significant factor. However, the counter-intuitiveness of this apparent influence of the laptop condition raises the question of interaction with another significant factor and the

possibility that scores should be adjusted by one or more factors before determining with any certitude the roles of the listed factors in the model. It is of particular interest therefore to consider analyzing students' performance when their scores are adjusted for their baseline reading levels and also to consider whether the implications of %FRPL-eligibility may differ in a poorer state compared to a rich one. The effects of these interactions are analyzed in the next section of this report.

Analysis showed that TOTAL scores (ORCA-Closed score + MC score) were significantly influenced by many factors. First and foremost, as expected based on results for both ORCA-Closed and Multiple Choice, ORM scores ($p < .0001$) accounted for the largest share of the Type III SS, 2353.1 out of 3723.8 (63.2%). As could also be predicted from the results for the two scores that make up the TOTAL, non-laptop students scored .6 points more than did laptop students ($p = .0045$); and Gender was also significant ($p < .0001$), with females scoring 1.8 points higher on average than males. Version was a significant factor as well ($p = .0236$), with a 0.9 range for mean TOTAL score from 16.5 to 17.4 across version pairs. Prior knowledge ($p < .0001$) accounting for 355.8 (9.6%) Type III SS and computer expertise ($p < .0001$) accounting for 351.3 (9.4%) were more important than %FRPL-eligible which accounted for ($p = .0211$) accounted for 43.9 Type III SS (1.2%). None of the order of the exam, number of computers in the home, or hours spent on the computer each day was a significant factor in this main effects model.

Lastly, Score differences (within student) that should reveal preferences or advantages between Multiple Choice and the ORCA-Closed were examined with few highly significant results. Gender was a significant factor ($p = .0002$), with males having a .5 point larger discrepancy between formats. Version of the test was also significant ($p < .0001$), with score differentials ranging between -3.6 and -2.6, always with Multiple Choice scores higher. Other factors contributed much less to the model although these were still statistically significant however the size of these effects was very small. Prior knowledge contributed 55.6 of the 407.1 (13.7%) Type III SS total score ($p = .0002$) Also hours spent on the computer was a significant factor ($p = .0494$) and accounted for 87.1 of the Type III SS score (21.4%). None of the laptop condition, ORM score, %FRPL, order of the exam, number of computers in the home, and expertise with the Internet contributed significantly to modeling the difference in difficulty between the internet ORCA-Closed and the non-internet Multiple Choice formats.

There were several salient results from these analyses of students' scores overall. First, ORM scores made up at least 50% of Type III SS across the board for measuring performance. However ORM had no role in explaining difference in relative ease/difficulty across formats. Second, females scored uniformly higher and their scores were more clustered than was true for males. However, it is generally the case with reading assessments that females score on average higher than males. The third important factor in modeling performance was computer expertise; the contribution to the model was consistently less never accounting for more than 10.9% of Type III SS but always significant, never accounting for more than 10.9% of Type III SS scores. The computer experience factor that contributed most to modeling the differential between formats was hours spent daily on the computer.

When scores are broken into subscores, different factors emerge as important for the four components; and Locate stands out as distinct from the other three. However, apart from Locate,

ORM is invariably significant. Prior knowledge is highly significant for both Evaluate and Synthesis. Topic Pair is highly significant for both Locate and Synthesis. Although Gender was found to be significant for each format and for Total as well, the only component for which Gender is significant is Synthesis. (Parenthetically, the R-square values for regression models for components are small. In part this is due to the small range of possible values (integers from 0 to 4) for each subscore. Nonetheless, these analyses give very good indications of the relative importance of the factors in each of these models.) As is indicated in other analyses and in the graphs of scores by Topic Pairs, the distinctions among the different versions are sufficient that further analyses of results should take into account or adjust for these differences in difficulty.

Generalized Linear Models (GLM): Analyses and Tables 7.1 – 7.9 (& 7.9a – 7.9d)

Factors in Models

FRPL:	% Eligible for Free/Reduced Price Lunch
ORM:	Offline Reading Measure
Laptop/Non-Laptop:	1= 1:1 Laptop (State 1); 2= No laptop (State 2)
Gender	1=Male; 2= Female
FORMATA:	Format order: 1= ORCA-Closed on Day 1; 2= MC on Day 2
FORMATB:	Format order: 1= MC on Day 1; 2= ORCA-Closed on Day 2
PKTM:	Score on Prior Knowledge Measure
Topic:	Version
Topic Pair:	AB: Day 1: Version A; Day 2: Version B
TOTAL score:	Closed score + MC score
Difference:	Closed score – MC score

Analysis 7.1 **ORCA-Closed:** ANALYSIS of ORCA-Closed Scores - GLM with Type III SS

SOURCE	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	3425.53714	380.61524	53.92	<.0001
Error	1271	8972.09908	7.05909		
Corrected Total	1280	12397.63622			

SOURCE	DF	Type III Sum of Squares	Mean Square	F Value	Pr > F
FRPL	1	45.314340	45.314340	6.42	0.0114
ORM	1	1818.183682	1818.183682	257.57	<.0001
Laptop/Non-Laptop	1	27.260956	27.260956	3.86	0.0496
Gender	1	466.717907	466.717907	66.12	<.0001
FORMATA	1	7.112278	7.112278	1.01	0.3157
PKTM	1	142.436806	142.436806	20.18	<.0001
Version/Topic	1	97.322469	97.322469	4.60	0.0033

R-Square	Coefficient of Variation	Root Mean Square Error	ORCA-Closed Mean
0.276306	39.06211	2.656894	6.801717

Analysis 7.2: **MC :** ANALYSIS of Multiple Choice Scores – G LM with Type III SS

SOURCE	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	4369.16335	485.46259	73.56	<.0001
Error	1271	8387.73127	6.59932		
Corrected Total	1280	12756.89461			

SOURCE	DF	Type III Sum of Squares	Mean Square	F Value	Pr > F
FRPL	1	35.142501	35.142501	5.33	0.0212
ORM	1	2298.973504	2298.973504	348.37	<.0001
Laptop/Non-Laptop	1	14.383186	14.383186	2.18	0.1401
Gender	1	144.643841	144.643841	21.92	<.0001
FORMATA	1	1.847850	1.847850	0.28	0.5968
PKTM	1	516.983037	516.983037	78.34	<.0001
Version/Topic	1	45.720268	45.720268	2.31	0.0748

R-Square	Coefficient of Variation	Root Mean Square Error	Multiple Choice Mean
0.342494	25.11853	2.568913	10.22717

Tables 7.3a and 7.3b: **Means and Standard Deviations by Format**

	CRCA-Closed			**Multiple Choice**		
	Sample Size	Mean	Std Dev	Sample Size	Mean	Std Dev
Male	612	6.18	3.17	612	9.89	3.30
Female	669	7.37	2.95	669	10.53	2.99
State 1	689	6.95	3.17	689	10.35	3.16
State 2	592	6.63	3.04	592	10.08	3.15

Version	**ORCA-Closed**			**Multiple Choice**		
	Sample Size	Mean	Std Dev	Sample Size	Mean	Std Dev
1	271	7.26	2.90	361	9.85	3.317
3	301	7.09	3.17	321	10.34	2.98
5	368	6.47	3.18	281	10.44	3.08
7	341	6.54	3.10	318	10.36	3.20

Analysis 7.4: **TOTAL:** ANALYSIS of TOTAL Scores - GLM with Type III SS

SOURCE	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	13	15444.82375	1188.06337	64.56	<.0001
Error	1267	23315.10755	18.40182		
Corrected Total	1280	38759.93130			

SOURCE	DF	Type III Sum of Squares	Mean Square	F Value	Pr > F
FRPL	1	178.594182	178.594182	9.71	0.0019
ORM	1	8157.943082	8157.943082	443.32	<.0001
Laptop/Non-Laptop	1	80.481426	80.481426	4.37	0.0367
Gender	1	1100.379560	1100.379560	59.80	<.0001
FORMATA	1	16.055926	16.055926	0.87	0.3504
PKTM	1	1114.596369	1114.596369	60.57	<.0001
Topic Pair	7	426.051111	60.864444	3.31	0.0017

R-Square	Coefficient of Variation	Root Mean Square Error	TOTAL Score Mean
0.398474	25.19093	4.289734	17.02888

Analysis 7.5: **Difference :** ANALYSIS of Difference Scores - GLM with Type III SS

SOURCE	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	13	503.12182	38.70168	4.44	<.0001
Error	1267	11046.00854	8.71824		
Corrected Total	1280	11549.13037			

SOURCE	DF	Type III Sum of Squares	Mean Square	F Value	Pr > F
FRPL	1	0.1789446	0.1789446	0.02	0.8861
ORM	1	17.9492515	17.9492515	2.06	0.1516
Laptop/Non-Laptop	1	0.9116861	0.9116861	0.10	0.7465
Gender	1	95.7237212	95.7237212	10.98	0.0009
FORMATA	1	0.5230731	0.5230731	0.06	0.8065
PKTM	1	100.3515948	100.3515948	11.51	0.0007
Topic Pair	7	218.5789694	31.2255671	3.58	0.0008

R-Square	Coefficient of Variation	Root Mean Square Error	Difference Score Mean
0.043564	86.19794	2.952666	-3.425449

Analysis 7.6: **Difference by Component :** ANALYSIS of Component Difference Scores: GLM with Type III SS

Analysis 7.6a: **Locate**

SOURCE	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	13	69.744286	5.364945	2.52	0.0020
Error	1267	2692.592950	2.125172		
Corrected Total	1280	2762.337237			

SOURCE	DF	Type III Sum of Squares	Mean Square	F Value	Pr > F
FRPL	1	0.00665487	0.00665487	0.00	0.9554
ORM	1	3.74974416	3.74974416	1.76	0.1843
Laptop/Non-Laptop	1	1.91558301	1.91558301	0.90	0.3426
Gender	1	3.57441576	3.57441576	1.68	0.1949
FORMATA	1	5.67569638	5.67569638	2.67	0.1025
PKTM	1	6.61177247	6.61177247	3.11	0.0780
Topic Pair	7	36.76647901	5.25235414	2.47	0.0160

R-Square	Coefficient of Variation	Root Mean Square Error	Locate Difference Score Mean
0.025248	241.2710	1.457797	-0.604215

Analysis 7.6b: **Evaluate**

SOURCE	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	13	41.394955	3.184227	2.06	0.0139
Error	1267	1955.637051	1.543518		
Corrected Total	1280	1997.032006			

SOURCE	DF	Type III Sum of Squares	Mean Square	F Value	Pr > F
FRPL	1	1.40607113	1.40607113	0.91	0.3400
ORM	1	7.25402205	7.25402205	4.70	0.0304
Laptop/Non-Laptop	1	0.05709054	0.05709054	0.04	0.8475
Gender	1	0.01934047	0.01934047	0.01	0.9109
FORMATA	1	2.83502762	2.83502762	1.84	0.1756
PKTM	1	9.03246692	9.03246692	5.85	0.0157
Topic Pair	7	8.92827828	1.27546833	0.83	0.5654

R-Square	Coefficient of Variation	Root Mean Square Error	Evaluate Difference Score Mean
0.020728	103.7480	1.242384	-1.197502

Analysis 7.6c: **Synthesize**

SOURCE	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	13	323.828104	24.909854	12.57	<.0001
Error	1267	2511.072755	1.981904		
Corrected Total	1280	2834.900859			

SOURCE	DF	Type III Sum of Squares	Mean Square	F Value	Pr > F
FRPL	1	1.3484600	1.3484600	0.68	0.4096
ORM	1	13.1279567	13.1279567	6.62	0.0102
Laptop/Non-Laptop	1	5.0685178	5.0685178	2.56	0.1100
Gender	1	77.9169527	77.9169527	39.31	<.0001
FORMATA	1	1.0314606	1.0314606	0.52	0.4708
PKTM	1	16.8559373	16.8559373	8.50	0.0036
Topic Pair	7	216.4017060	30.9145294	15.60	<.0001

R-Square	Coefficient of Variation	Root Mean Square Error	Synthesize Difference Score Mean
0.114229	862.8676	1.407801	-0.163154

Analysis 7.6d: **Communicate**

SOURCE	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	13	43.838462	3.372189	1.95	0.0216
Error	1267	2190.420710	1.728825		
Corrected Total	1280	2234.259173			

SOURCE	DF	Type III Sum of Squares	Mean Square	F Value	Pr > F
FRPL	1	0.10042208	0.10042208	0.06	0.8096
ORM	1	10.43388930	10.43388930	6.04	0.0142
Laptop/Non-Laptop	1	0.02292339	0.02292339	0.01	0.9083
Gender	1	1.15109410	1.15109410	0.67	0.4147
FORMATA	1	5.41623033	5.41623033	3.13	0.0770
PKTM	1	0.11237448	0.11237448	0.07	0.7988
Topic Pair	7	22.17698632	3.16814090	1.83	0.0774

R-Square	Coefficient of Variation	Root Mean Square Error	Communicate Difference Score Mean
0.019621	90.02244	1.314848	-1.460578

Table7.7: **Means and Standard Deviations for TOTAL and Difference Scores by Gender**

Gender	Male			Female		
	Sample Size	Mean	Std Dev	Sample Size	Mean	Std Dev
TOTAL	612	16.07	5.72	669	17.91	5.14
Difference	612	-3.72	3.02	669	-3.16	2.96
D-Locate	612	-0.66	1.47	669	-0.55	1.47
D-Evaluate	612	-1.21	1.22	669	-1.19	1.28
D-Synthesize	612	-0.42	1.48	669	0.07	1.46
D-Communicate	612	-1.43	1.33	669	-1.49	1.32

Table7.8: **Means and Standard Deviations for TOTAL and Difference Scores by 1:1 Laptop/Non-Laptop (by State)**

Laptop/ Non-Laptop	1:1 Laptop (State 1)			Non-Laptop(State 2)		
	Sample Size	Mean	Std Dev	Sample Size	Mean	Std Dev
TOTAL	689	17.31	5.63	592	16.71	5.33
Difference	689	-3.40	2.88	592	-3.45	3.14
D-Locate	689	-0.64	1.46	592	-0.56	1.48
D-Evaluate	689	-1.19	1.22	592	-1.20	1.28
D-Synthesize	689	-0.10	1.50	592	-0.23	1.47
D-Communicate	689	-1.46	1.28	592	-1.46	1.36

Tables 7.9a, 7.9b, 7.9c &7.9d: **Means and Standard Deviations for TOTAL and Difference Scores by Topic Pairs**

Topic Pair 1&3	1 , 3			3 , 1		
	Sample Size	Mean	Std Dev	Sample Size	Mean	Std Dev
TOTAL	133	17.22	5.45	161	16.91	6.04
Difference	133	-2.98	2.93	161	-2.63	3.00
D-Locate	133	-0.63	1.29	161	-0.27	1.42
D-Evaluate	133	-1.07	1.33	161	-1.10	1.36
D-Synthesize	133	0.19	1.54	161	0.38	1.50
D-Communicate	133	-1.47	1.38	161	-1.63	1.32

Topic Pair 1&7	1,7			7,1		
	Sample Size	Mean	Std Dev	Sample Size	Mean	Std Dev
TOTAL	138	18.37	4.85	200	16.50	5.68
Difference	138	-3.57	2.62	200	-3.32	2.97
D-Locate	138	-0.84	1.57	200	-0.84	1.53
D-Evaluate	138	-1.31	1.16	200	-1.15	1.18
D-Synthesize	138	0.28	1.26	200	0.11	1.41
D-Communicate	138	-1.70	1.26	200	-1.44	1.19

Topic Pair 5&3	5 , 3			3 , 5		
	Sample Size	Mean	Std Dev	Sample Size	Mean	Std Dev
TOTAL	188	17.40	5.26	140	18.13	5.20
Difference	188	-3.63	3.18	140	-4.06	3.11
D-Locate	188	-0.52	1.45	140	-0.61	1.65
D-Evaluate	188	-1.13	1.14	140	-1.34	1.35
D-Synthesize	188	-0.60	1.53	140	-0.56	1.45
D-Communicate	188	-1.39	1.38	140	-1.54	1.42

Topic Pair 5&7	5,7			7,5		
	Sample Size	Mean	Std Dev	Sample Size	Mean	Std Dev
TOTAL	180	15.92	5.70	141	16.24	5.21
Difference	180	-3.85	3.03	141	-3.33	2.88
D-Locate	180	-0.66	1.37	141	-0.43	1.36
D-Evaluate	180	-1.25	1.25	141	-1.27	1.25
D-Synthesize	180	-0.64	1.36	141	-0.35	1.41
D-Communicate	180	-1.30	1.27	141	-1.28	1.33

8. *Question: After accounting for general reading ability (ORM) and version pair of the test, which factors drive performance from among test attributes, school attributes, student attributes, and student computer familiarity?*

Because basic reading ability is not equivalently distributed either between males and or between states or according to other factors, understanding the influence of these factors rather than their reflection of baseline reading ability requires using adjusted scores. In this case, standardized (IRT) scores were adjusted using regression and the two factors that otherwise are seriously confounded with all the rest: the feature of the particular test: Version Pair and baseline reading ability measured by ORM.

As before to investigate factors that influence scores, we looked at scores for two Internet setups in four ways: Multiple Choice, ORCA-Closed, total scores, and the difference between Multiple Choice and Closed scores for each student. Of these, the first three look at factors affecting scores; and the differences investigate specific discrepancies between Multiple Choice and ORCA-Closed. Specific data for all the results summarized below are documented in the Analyses and Tables that follow. The specific factors investigated within the models themselves are:

laptop (state 1) or no laptop (state 2) in school
gender
offline reading measurement (ORM)
socioeconomic status (FRPL)
order of test
prior knowledge (PKTM)
relative difficulty of version, adjustment for format
personal computer at home
number of computers in the home
hours of computer use per day
expertise of computer use

Note that Version and version pairs of the test (that was included in models in Section 7), were not included in this analysis because the scores in the current section were adjusted for Version Pair.

For the Multiple Choice scores, there was a range of results. Gender was still a significant factor, ($p = .0012$), with females scoring .55 points higher than did males. However, prior knowledge was more important for scores ($p < .0001$), contributing 216.2 points of the Type III SS score (40.0%). Expertise with computers was significant as well ($p < .0001$), accounting for 169.8 points (31.4%). The laptop condition, as a factor, was marginally significant ($p = .0623$), with non-laptop students scoring .33 point higher, on average, than laptop students. FRPL was similarly marginal ($p = .0695$), accounting for 17.2 points of the Type III SS score (3.2%). The non-significant factors for students' scores were the order of the test, personal computer at home, number of computers in the home, and hours spent on the computer daily.

For ORCA-Closed there were many important factors for student scoring. The laptop condition (state) was significant ($p = .0076$), with non-laptop students again scoring .33 points higher than did laptop students. Gender was also an important factor ($p < .0001$): females scored on average .92 points higher than males did, a larger differential than for Multiple Choice. FRPL was

significant (p = .0201), accounting for 22.5 of the 468.9 Type III SS score (4.8%); prior knowledge (p = .0011) contributed 44.4 points (9.5%); hours (p = .0189) accounted for 101.5 (21.6%); and computer expertise (p = .0015) added 64.5 points (13.8%). Order of the tests, personal computer at home and number of computers in the home were both non-significant in terms of impact on students' scores.

The same factors were significant for TOTAL scores as for Multiple Choice. For the laptop condition's significant effect (p = .0093), non-laptop students scored .66 point higher than laptop students; and gender was also important (p < .0001), with females scoring on average 1.46 points higher than males. FRPL was a significant factor as well (p = .0125), contributing 79.5 points of the total Type III SS score of 1658.5 (4.8%). Prior knowledge and expertise were also quite significant (p < .0001) – it accounted for 461.3 points (27.8%). Computer expertise contributed 432 points (26.0%). Once again, test order, personal computer, number of computers in the home, and hours spent on the computer daily were not significant as factors.

Looking at the Differences in scores between formats, few results were significant. Gender was significant (p = .0122), with females performing better on ORCA-Closed format (average score on ORCA-Closed was 0.19 higher than on ORCA-MC) while males performed better on multiple choice format (scores on ORCA-Closed was 0.18 higher than on ORCA-MC). Prior knowledge (p = .0013) contributed 62.4 of 244.3 points of the Type III SS score (25.5%). The non-significant factors were laptop condition, FRPL, test order, number of computers in the home, hours spent on the computer daily, and computer expertise.

There were several intriguing results culled from the data. First and foremost, gender and prior knowledge were significant factors across formats for all topic combinations. For gender, this comprised a consistent significant impact for all conditions (1.46 for TOTAL of Multiple Choice and ORCA-Closed combined). Prior knowledge was also consistently important, however the magnitude of its impact was differed greatly as the percentage of the Type III SS scores accounted for ranged from 9.5% to 40%. At the other end of the spectrum, the order of the test, number of computers in the home, and hours spent daily on the computer were consistently non-significant as factors. This indicates that students did not gain or lose anything between tests, and that their use of computers outside of school had little impact on their use of computers for the ORCA testing. In sum, ORCA scores were affected by what students knew going into the exam, not how they accessed new information; and females were simply higher scoring on average.

Generalized Linear Models (GLM): Analyses and Tables 8.1 – 8.6

Response Variables: Standardizes (IRT) Scores adjusted for Offline Reading Score & Version Difficulty
- TOTAL score: Closed score + MC score
- Difference: Closed score – MC score

Factors in Models
- FRPL: % Eligible for Free/Reduced Price Lunch
- Laptop/Non-Laptop: 1= 1:1 Laptop (State 1); 2= No laptop (State 2)
- Gender: 1=Male; 2= Female
- FORMATA: Format order: 1= ORCA-Closed on Day 1; 2= MC on Day 2
- FORMATB: Format order: 1= MC on Day 1; 2= ORCA-Closed on Day 2
- PKTM: Score on Prior Knowledge Measure
- Relative Difficulty: Relative Version Difficulty Adjustment between ORCA-Closed and MC
- Personal Computer: Dedicated Personal Home Computer
- Home Computer: Number of Computers at Home
- Computer Time: Hours Daily at Computer or Online
- Computer Expertise: Self-assessment of Expertise

Analysis 8.1 **ORCA-Closed:** ANALYSIS of Adjusted Scores - GLM with Type III SS

SOURCE	DF	Type III Sum of Squares	Mean Square	F Value	Pr > F
FRPL	1	22.5421040	22.5421040	5.41	0.0201
Laptop/Non-Laptop	1	29.7771778	29.7771778	7.15	0.0076
Gender	1	183.7595521	183.7595521	44.14	<0.0001
FORMATA	1	7.7075591	7.7075591	1.85	0.1739
PKTM	1	44.4013801	44.4013801	10.67	0.0011
Relative Difficulty	3	4.3801978	1.4600659	0.35	0.7887
Personal Computer	1	2.5621160	2.5621160	0.62	0.4329
Home Computers	10	14.6644613	1.4664461	0.35	0.9661
Computer Time	12	101.4928346	8.4577362	2.03	0.0189
Computer Expertise	3	64.4903417	21. 4967806	5.16	0.0015

R-Square	Coefficient of Variation	Root Mean Square Error	ORCA-Closed Adjusted Score Mean
0.107277	2698173	2.040411	

Analysis 8.2: **MC :** ANALYSIS of Multiple Choice Scores – G LM with Type III SS

SOURCE	DF	Type III Sum of Squares	Mean Square	F Value	Pr > F
FRPL	1	17.1506356	17.1506356	3.30	0.0695
Laptop/Non-Laptop	1	18.0918862	18.0918862	3.48	0.0623
Gender	1	54.4969688	54.4969688	10.49	0.0012
FORMATB	1	4.4983781	4.4983781	0.87	0.3523
PKTM	1	216.1835788	216.1835788	41.61	<0.0001
Relative Difficulty	3	2.4403939	0.8134646	0.16	0.9255
Personal Computer	1	2.2109137	2.2109137	0.43	0.5143
Home Computers	10	39.1462665	3.9146266	0.75	0.6740
Computer Time	12	21.6582788	1.8048566	0.35	0.9800
Computer Expertise	3	169.7504272	56.5834757	10.89	<0.0001

R-Square	Coefficient of Variation	Root Mean Square Error	Multiple Choice Mean
0.105995	21943.29	2.279279	-0.010387

Analysis 8.3: **TOTAL :** ANALYSIS of Adjusted TOTAL Scores – G LM with Type III SS

SOURCE	DF	Type III Sum of Squares	Mean Square	F Value	Pr > F
FRPL	1	79.5071925	79.5071925	6.25	0.0125
Laptop/Non-Laptop	1	86.1921339	86.1921339	6.78	0.0093
Gender	1	441.9042660	441.9042660	34.75	<0.0001
FORMATB	1	24.7471365	24.7471365	1.95	0.1633
PKTM	1	461.3329750	461.3329750	36.28	<0.0001
Personal Computer	1	8.5930356	8.5930356	0.68	0.4112
Home Computers	10	46.8497298	4.6849730	0.37	0.9602
Computer Time	12	132.8887923	11.0740660	0.87	0.5767
Computer Expertise	3	432.0148168	144.0049389	11.32	<0.0001

R-Square	Coefficient of Variation	Root Mean Square Error	Adjusted TOTAL Score Mean
0132544	34582.02	3.565929	-0.010312

Analysis 8.4: **Difference :** ANALYSIS of Differences of Adjusted Scores – G LM with Type III SS

SOURCE	DF	Type III Sum of Squares	Mean Square	F Value	Pr > F
FRPL	1	0.4243590	0.4243590	0.07	0.7897
Laptop/Non-Laptop	1	1.2460913	1.2460913	0.21	0.6477
Gender	1	37.5812196	37.5812196	6.30	0.0123
FORMATB	1	0.3870015	0.3870015	0.03	0.7990
PKTM	1	62.3565205	62.3565205	10.45	0.0013
Personal Computer	1	0.0369678	0.0369678	0.01	0.9373
Home Computers	10	59.4084343	5.9408434	1.00	0.4448
Computer Time	12	111.8294015	9.3191168	1.56	0.0966
Computer Expertise	3	33.5225485	11.1741828	1.87	0.1323

R-Square	Coefficient of Variation	Root Mean Square Error	Mean Difference of Adjusted Scores
0.043578	23343,63	2.442384	0.010463

Table 8.5: **Medians, Means and Standard Deviations for Adjusted Scores by Gender**

	Male					**Female**				
	Sample Size	Median	Mean	Std Dev	Std Err	Sample Size	Median	Mean	Std Dev	Std Err
ORCA-Closed	579	-0.442	-0.484	2.166	0.090	644	0.510	0.436	1.999	0.079
Multiple Choice	579	-0.240	-0.298	2.406	0.100	644	0.312	0.248	2.322	0.091
TOTAL	579	-0.789	-0.782	3,851	0.160	644	0.763	0.683	3.578	0.141
Difference	579	-0.127	-0.187	2.477	0.103	644	0.141	0.188	2.444	0.096

Table 8.6: **Medians, Means and Standard Deviations for Adjusted Scores by 1:1 Laptop/ Non-Laptop (by State)**

	1:1 Laptop (State 1)					**Non-Laptop (State 2)**				
	Sample Size	Median	Mean	Std Dev	Std Err	Sample Size	Median	Mean	Std Dev	Std Err
ORCA-Closed	651	0.298	0.154	2.137	0.084	572	-0.141	-0.176	2.108	0.088
Multiple Choice	651	0.229	0.145	2.305	0.090	572	-0.078	-0.187	2.446	0.102
TOTAL	651	0.374	0.299	3.771	0.148	572	-0.264	-0.363	3.762	0.157
Difference	651	-0.058	0.010	2.354	0.092	572	0.038	0.011	2.589	0.108

8a: *Question Supplement: Do interpretations change when Models for Adjusted Scores also account for state differences in computer experiences?*

Of specific interest is the extent to which students' computer exposure and experience outside school might have a different effect for students who are in classrooms with 1:1 laptops compared to classrooms without laptops. This kind of effect could show in a model as an interaction between the laptop/non-laptop factor and the factors indicating out-of-school computer use. Adjusted scores (Standardized (IRT) scores adjusted for ORM and for Version Pair) are used in the model that has three additional interaction terms.

Investigating of factors that influence adjusted scores, we looked at scores for two Internet setups in four ways: Multiple Choice, ORCA-Closed, total scores, and the difference between Multiple Choice and Closed scores for each student. Of these, the first three look at factors affecting scores; and the differences investigate specific discrepancies between Multiple Choice and ORCA-Closed. Specific data for all the below results can be found in the analyses and tables in that follow. The specific factors investigated within the models themselves are:

laptop (state 1) or no laptop (state 2) in school
gender
socioeconomic status (FRPL)
order of test
prior knowledge (PKTM)
relative difficulty of version, adjustment for format
personal computer at home
number of computers in the home
hours of computer use per day
expertise of computer use

In addition, we looked at interactions between the laptop or not laptop condition and other factors, which created factors of:

laptop (state) -computers
laptop (state)-hours
laptop (state)-expertise

As in the previous section, the adjusted scores accounted for baseline reading (ORM) and Version Pair, so these do not appear in the list of modeled factors.

For the ORCA-Closed several factors were significant. Gender was significant ($p < .0001$), with males scoring .92 points *higher* than females once adjusted for ORM and Version Pair. FRPL was also significant ($p = .0364$), accounting for 18.4 points of the total 494.6-point Type III SS score (3.7%). Prior knowledge was important as well ($p = .0009$), accounting for 46.7 points (9.4%). Lastly, hours spent on the computer was significant as a factor ($p = .0266$), making up 97.5 points (19.7%); and expertise with computers added 53.8 points (10.9%). The laptop/non-laptop condition was marginally significant ($p = .0575$), with non-laptop students scoring .33 points more

than did laptop students. Test order, personal computer and number of computers in the home were all non-significant as factors. In the case of the interactions, none were significant as factors.

The Multiple Choice setup had few significant factors, and one marginally significant interaction factor. Gender was significant ($p = .0018$), with males scoring .55 points *higher* than females for adjusted scores. Prior knowledge was also important, accounting for 227.3 points of the 708.8 points making up the Type III SS score (32.1%); and computer expertise was likewise ($p < .0001$), garnering 188.3 points (26.6%). Laptop condition, FRPL, test order, number of computers in the home, and hours spent daily on the computer were unimportant as factors. For the interactions, state-expertise was marginally significant ($p = .0638$), with 37.7 points accounted for (5.3%). For both state-computers and state-hours, though, there were no significant results.

TOTAL adjusted score results were similar to those of the ORCA-Closed condition. The laptop condition was significant ($p = .0414$), with non-laptop students scoring .66 points higher than did laptop students when adjusted; and gender was also significant ($p < .0001$), with males scoring 1.46 points *higher* than did females. FRPL was important was well ($p = .0265$), accounting for 62.9 of 1927 Type III SS points (3.3%). Prior knowledge ($p < .0001$) added 487.6 points (25.3%); and computer expertise ($p < .0001$) accounted for 433.8 (22.5%). There were also non-significant results from the test order, number of computers in the home, and hours spent on the computer. For the interactions, all three were non-significant as factors.

The Difference scores between formats followed a similar pattern to the results for the Multiple Choice format. Gender was significant as a factor ($p = .0105$), with males having .01 points greater difference than did females. Prior knowledge was important ($p = .0009$), accounting for 65.3 points of the 479-point Type II SS total score (13.6%); as was computer expertise ($p = .0486$), garnering 46.9 points (9.8%). The laptop condition, FRPL, test order, number of computers in the home, and hours spent on the computer daily were all non-significant factors for students' scores. The interactions were no different as the latter group, with all being non-significant factors.

The results of the adjusted score analyses produced some interesting overall results. The first is the consistent significance of several factors: gender, prior knowledge, and computer expertise factors for all four score analyses. This implies that these three factors are at the very least among the most important factors for student performance. For the Multiple Choice and Difference scores, these three are the only factors that are truly significant (the state-expertise being marginal). So since ORCA-Closed and TOTAL scores both have the additional significant factors of laptop condition, FRPL and number of hours daily (computer daily use), these three added factors putatively influence performance skills tested by ORCA-Closed but not by Multiple Choice. It is conjectured that %FRPL-eligible is likely a confounded factor with computer exposure/experience factors as well. That being said, for the always-present factors, FRPL was consistently of low impact, as it accounted for under 4% of the Type III SS scores; and both prior knowledge and expertise were of varying impact on the different formats.

Generalized Linear Models (GLM) with Interactions: Analyses 8A.1 – 8A.4

Response Variables: Standardized (IRT) Scores adjusted for Offline Reading Score & Version Difficulty
- TOTAL score: Closed score + MC score
- Difference: Closed score – MC score

Factors in Models
- FRPL: % Eligible for Free/Reduced Price Lunch
- Laptop/Non-Laptop: 1= 1:1 Laptop (State 1); 2= No laptop (State 2)
- Gender: 1=Male; 2= Female
- FORMATA: Format order: 1= ORCA-Closed on Day 1; 2= MC on Day 2
- FORMATB: Format order: 1= MC on Day 1; 2= ORCA-Closed on Day 2
- PKTM: Score on Prior Knowledge Measure
- Relative Difficulty: Relative Version Difficulty Adjustment between ORCA-Closed and MC
- Personal Computer: Dedicated Personal Home Computer
- Home Computer: Number of Computers at Home
- Computer Time: Hours Daily at Computer or Online
- Computer Expertise: Self-assessment of Expertise
- 1:1 Laptop x Personal Computer: Differential by Laptop in Effect of Personal Computer
- 1:1 Laptop x Home Computer: Differential by Laptop in Effect of Number of Computers
- 1:1 Laptop x Computer Time: Differential by Laptop in Effect of Computer Time
- 1:1 Laptop x Computer Expertise: Differential by Laptop in Effect of Expertise

Analysis 8A.1 **ORCA-Closed:** ANALYSIS of ORCA-Closed Adjusted Scores - GLM with Type III SS

SOURCE	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	60	671.570512	11.192842	2.67	<0.0001
Error	1162	4868.754142	4.189978		
Corrected Total	1222	5540.324655			

SOURCE	DF	Type III Sum of Squares	Mean Square	F Value	Pr > F
FRPL	1	18,3902322	18,3902322	4.39	0.0364
Laptop/Non-Laptop	1	15.1511598	15.1511598	3.62	0.0575
Gender	1	179.1314027	179.1314027	42.75	<0.0001
FORMATA	1	6,0696312	6,0696312	1.45	0.2290
PKTM	1	46.6767072	46.6767072	11.14	0.0009
Relative Difficulty	3	5.0369419	1.6789806	0.40	0.9414
Personal Computer	1	3.9593112	3.9593112	0.94	0.3312
Home Computers	10	17.2649767	1.7264977	0.41	0.9414
Computer Time	12	97.47333252	8,1227771	1.94	0.0266
Computer Expertise	3	53,7533898	17.9177966	4.28	0.0052

1:1 Laptop x Personal Computer	1	6.0610630	6.0610630	1.45	0.2293
1:1 Laptop x Home Computers	10	24,6735269	2.4673527	0.59	0.8241
1:1 Laptop Computer Time	12	34.3845458	2.8653788	0.68	0.7683
Computer Expertise	3	11.3805279	0.4601760	0.11	0.9544

R-Square	Coefficient of Variation	Root Mean Square Error	ORCA-Closed Mean
0.121215	2706811	2.046944	0.00007660

Analysis 8A.2: **MC :** ANALYSIS of Multiple Choice Adjusted Scores – G LM with Type III SS

SOURCE	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	60	894.612810	14.910214	2.88	<0.0001
Error	1162	6008.916407	5.171185		
Corrected Total	1222	6903.529217			

SOURCE	DF	Type III Sum of Squares	Mean Square	F Value	Pr > F
FRPL	1	12.9672942	12.9672942	2.51	0.1136
Laptop/Non-Laptop	1	13.4572371	13.4572371	2.60	0.1070
Gender	1	50.6196510	50.6196510	9.79	0.0018
FORMATA	1	5.1784033	5.1784033	1.00	0.3172
PKTM	1	227.2874256	227.2874256	43.95	<0.0001
Relative Difficulty	3	1.2927638	0.4309213	0.08	0.9691
Personal Computer	1	0.9609763	0.9609763	0.19	0.6665
Home Computers	10	28.1194498	2.8119450	0.54	0.8596
Computer Time	12	19.4624067	1.6218672	0.31	0.9872
Computer Expertise	3	188.3278609	62.7759536	12.14	<0.0001
1:1 Laptop x Personal Computer	1	3.0634979	3.0634979	0.59	0.4416
1:1 Laptop x Home Computers	10	68.2345824	6.8234582	1.32	0.2145
1:1 Laptop Computer Time	12	57.0452846	4.7537737	0.092	0.5266
Computer Expertise	3	37.6938662	12.5646221	2.43	0.0638

R-Square	Coefficient of Variation	Root Mean Square Error	ORCA-Closed Mean
0.129588	21892.70	2.274024	-0.010387

Analysis 8A.3: **TOTAL :** ANALYSIS of Total Adjusted Scores – G LM with Type III SS

SOURCE	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	57	2609.26754	45.77662	3.59	<0.0001
Error	1165	14849.34878	12.74622		
Corrected Total	1222	17458,61632			

SOURCE	DF	Type III Sum of Squares	Mean Square	F Value	Pr > F
FRPL	1	62.8702579	62.8702579	4.93	0.0265
Laptop/Non-Laptop	1	53.1500016	53.1500016	4.17	0.414
Gender	1	423.1901988	423.1901988	33.20	<0.0001
FORMATA	1	23.6891843	23.6891843	1.86	0.1731
PKTM	1	487.5803632	487.5803632	38.25	<0.0001
Personal Computer	1	7.9909816	7.9909816	0.63	0.4286
Home Computers	10	35.4137398	3.5413740	0.28	0.9860
Computer Time	12	128.3708730	10.6975727	0.84	0.6097
Computer Expertise	3	433.8176796	144.6058932	11.35	<0.0001
1:1 Laptop x Personal Computer	1	0.2960819	0.2960819	0.02	0.8789
1:1 Laptop x Home Computers	10	123.8278012	12.3827801	0.97	0.4665
1:1 Laptop Computer Time	12	102.3927371	8.5327281	0.67	0.7820
1:1 Laptop x Computer Expertise	3	52.6380841	15.5460280	1.38	0.2484

R-Square	Coefficient of Variation	Root Mean Square Error	ORCA-Closed Mean
0.149454	34623.30	3.570185	-0.010312

Analysis 8A.4: **Difference :** ANALYSIS of Difference of Adjusted Scores – G LM with Type III SS

SOURCE	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	57	510.439689	8.988082	1.51	0.0097
Error	1164	6918.651733	5.938757		
Corrected Total	1222	7429.091422			

SOURCE	DF	Type III Sum of Squares	Mean Square	F Value	Pr > F
FRPL	1	0.5201184	0.5201184	0.09	0.7673
Laptop/Non-Laptop	1	0.0076780	0.0076780	0.00	0.9713
Gender	1	39.0079460	39.0079460	6.57	0.0105
FORMATA	1	0.0151575	0.0151575	0,00	0.9598
PKTM	1	65.3188565	65.3188565	11.00	0.0009
Personal Computer	1	1.0021898	1.0021898	0.17	0.6813

Home Computers	10	54.1230365	5.4123037	0.91	0.5218
Computer Time	12	103.1506365	8,5958864	1.45	0.1381
Computer Expertise	3	46.9127187	15.6375729	2.63	0.0486
1:1 Laptop x Personal Computer	1	16.6415406	16.6415406	2.80	0.944
1:1 Laptop x Home Computers	10	63.3294744	6.3329474	1.07	0.3853
1:1 Laptop Computer Time	12	80.3516494	6.6959708	1.13	0.3331
1:1 Laptop x Computer Expertise	3	26.3245157	8.7748386	1.48	0.2190

R-Square	Coefficient of Variation	Root Mean Square Error	ORCA-Closed Mean
0.06708	23291.73	2.436956	0.010463

8b: *Question Supplement: Do interpretations change when Models for Adjusted Scores also account for state differences in economics?*

Economic differences between states also raise the question of differential impact of FRPL in each state. To address this question, standardized (IRT) scores were modeled with including factors listed in the previous section but adding a state-FRPL interaction. The state-computer variable interactions were retained since computer expertise and daily exposure might affect ORCA scores differently for students in a laptop or non-laptop state.

For the ORCA-Closed, of all interaction terms added, only state-FRPL was significant ($p = .0386$). The inclusion of the interaction term increased the R^2 term from .28 to .33.

For the Multiple Choice scores, there was similarly only a marginally significant term for state-FRPL ($p = .0862$). However inclusion of this interaction in the model increased the R^2 from .34 to .39.

For TOTAL performance, state-FRPL was the only significant factor among the added interactions. Although marginally significant ($p = .0557$) including this term increased the R^2 from .40 to .45.

Adding interaction terms did not improve the model fit for the Difference in scores between ORCA-Closed and Multiple Choice.

Similarly adding interaction terms for state-computer variables yielded no benefit in modeling any of ORCA-Closed, Multiple Choice, TOTAL or Difference.

Generalized Linear Models (GLM) with Interactions: Analyses 8B.1 – 8B.4

Response Variable: Standardized (IRT) Scores adjusted for Version Difficulty)
TOTAL Standardized Score: ORCA-Closed score + MC score
Difference: ORCA-Closed score – MC score

Factors in Models

- FRPL: % Eligible for Free/Reduced Price Lunch
- ORM: Offline Reading Measure Score
- Laptop/Non-Laptop: 1= 1:1 Laptop (State 1); 2= No laptop (State 2)
- Gender: 1=Male; 2= Female
- FORMATA: Format order: 1= ORCA-Closed on Day 1; 2= MC on Day 2
- FORMATB: Format order: 1= MC on Day 1; 2= ORCA-Closed on Day 2
- PKTM: Score on Prior Knowledge Measure
- Version: Version (ORCA-Closed or Multiple Choice, as appropriate)
- Topic Pairs: Ordered Combination of Day 1 and Day2 Topics
- Personal Computer: Dedicated Personal Home Computer
- Home Computer: Number of Computers at Home
- Computer Time: Hours Daily at Computer or Online
- Computer Expertise: Self-assessment of Expertise
- State I1:1 Laptop) x FRPL: Differential by State (1:1 Laptop) in Effect of % FRPL
- 1:1 Laptop x Personal Computer: Differential by Laptop in Effect of Personal Computer
- 1:1 Laptop x Home Computer: Differential by Laptop in Effect of Number of Computers
- 1:1 Laptop x Computer Time: Differential by Laptop in Effect of Computer Time
- 1:1 Laptop x Computer Expertise: Differential by Laptop in Effect of Expertise

Analysis 8B.1 **ORCA-Closed:** ANALYSIS of ORCA-Closed Standardized Scores - GLM with Type III SS

SOURCE	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	62	2362.576033	38,106065	9.19	<0.0001
Error	1160	4808.123665	4.144934		
Corrected Total	1222	7170.699698			

SOURCE	DF	Type III Sum of Squares	Mean Square	F Value	Pr > F
FRPL	1	36.4798754	36.4798754	8.80	0.0031
ORM	1	730.6501927	730.6501927	176.28	<0.0001
Laptop/Non-Laptop	1	0.1477331	0.1477331	0.04	0.8503
Gender	1	177.3497178	177.3497178	42.79	<0.0001
FORMATA	1	5.4582012	5.4582012	1.32	0.2514
PKTM	1	72.6409414	72.6409414	17.53	<0.0001
Version (ORCA-Closed)	3	88.8825124	29,6275041	7.15	<0.0001
Personal Computer	1	3.2491990	3.2491990	0.78	0.3761
Home Computers	10	15.4730761	1.5473076	0.37	0.9583
Computer Time	12	95.1601335	7.9300111	1.91	0.0292
Computer Expertise	3	77.0791275	25.6930425	6.20	0.0004

State (1:1 Laptop) x FRPL	1	17.7819769	17.7819769	4.29	0.0386
1:1 Laptop x Personal Computer	1	3.2491990	3.2491990	0.78	0.3761
1:1 Laptop x Home Computers	10	22.6737705	2.2673771	0.55	0.8572
1:1 Laptop Computer Time	12	39,9304339	3.3275362	0.80	0.6479
1:1 Laptop x Computer Expertise	3	1.0141456	0.3380485	0.08	0.9701

R-Square	Coefficient of Variation	Root Mean Square Error	ORCA-Closed Mean
0.329476	1054.823	2.035911	0.193010

Analysis 8B.2: **MC :** ANALYSIS of Multiple Choice Standardized Scores – G LM with Type III SS

SOURCE	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	62	3862.691320	62,301473	12.17	<0.0001
Error	1160	5939.517900	5.120274		
Corrected Total	1222	9802.209220			

SOURCE	DF	Type III Sum of Squares	Mean Square	F Value	Pr > F
FRPL	1	19.651445	19.651445	3.84	0.0503
ORM	1	1279.597818	1279.597818	249.91	<0.0001
Laptop/Non-Laptop	1	0.459312	0.459312	0.09	0.7646
Gender	1	52.348532	52.348532	19,22	0.0014
FORMATA	1	3,493982	3,493982	0.68	0.4089
PKTM	1	318.546355	318.546355	62.21	<0.0001
Version (Multiple Choice)	3	45.463637	15.154546	2.96	0.0314
Personal Computer	1	0.043339	0.043339	0.01	0.9267
Home Computers	10	31.365719	3.136572	0.61	0.8042
Computer Time	12	17.421248	1.451771	0.28	0.9919
Computer Expertise	3	245.490107	81.830036	25.98	<0.0001
State (1:1 Laptop) x FRPL	1	15.100312	15.100312	2.95	0.0862
1:1 Laptop x Personal Computer	1	2.218811	2.218811	0.43	0.5105
1:1 Laptop x Home Computers	10	60.702793	6.070279	1.19	0.2962
1:1 Laptop Computer Time	12	61.866322	5.155527	1.01	0.4399
1:1 Laptop x Computer Expertise	3	23.789901	7.929967	1.55	0.2002

R-Square	Coefficient of Variation	Root Mean Square Error	Multiple Choice Mean
0.394063	1617.468	2.262802	-0.139898

Analysis 8B.3: **TOTAL :** ANALYSIS of Total Standardized Scores – G LM with Type III SS

SOURCE	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	66	11659.37044	176.65713	14.09	<0.0001
Error	1156	14498.38211			
Corrected Total	1222	26157.75255			

SOURCE	DF	Type III Sum of Squares	Mean Square	F Value	Pr > F
FRPL	1	117.798966	117.798966	9.39	0.0022
ORM	1	3910.397614	3910.397614	311.79	<0.0001
Laptop/Non-Laptop	1	0.000206	0.000206	0.00	0.9968
Gender	1	407.048403	407.048403	32.46	<0.0001
FORMATA	1	19.181102	19.181102	1.53	0.2165
PKTM	1	651.091553	651.091553	51.91	<0.0001
Topic Pair	7	190.881097	27.268728	2.17	0.0341
Personal Computer	1	4.475739	4.475739	0.36	0.5504
Home Computers	10	40.290101	4.029101	0.32	0.9758
Computer Time	12	116.400228	9.700019	0.77	0.6785
Computer Expertise	3	571,781463	190.593821	15.20	<0.0001
State (1:1 Laptop) x FRPL	1	46.019883	46.019883	3.67	0.557
1:1 Laptop x Personal Computer	1	2.793321	2.793321	0.22	0.6371
1:1 Laptop x Home Computers	10	104. 484598	20.448460	0.83	0.5967
1:1 Laptop Computer Time	12	125.529428	20.460786	0.83	0.6152
1:1 Laptop x Computer Expertise	3	26.548929	8,849643	0.71	0.5487

R-Square	Coefficient of Variation	Root Mean Square Error	TOTAL Standardized Score Mean
0.445733	6667.884	3.541448	0.053112

Analysis 8B.4: **Difference :** ANALYSIS of Difference of Adjusted Scores – G LM with Type III SS

SOURCE	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	99	885.001117	13.409108	2.25	<0.0001
Error	1156	6903.064170	5.971509		
Corrected Total	1222	7788.065287			

SOURCE	DF	Type III Sum of Squares	Mean Square	F Value	Pr > F
FRPL	1	0.7313034	0.7313034	0.12	0.7264
ORM	1	68.3784114	68.3784114	11.45	0.0007
Laptop/Non-Laptop	1	0.1386998	0.1386998	0.02	0.8789
Gender	1	39.1155660	39.1155660	6.55	0.0106
FORMATA	1	0.0149181	0.0149181	0.00	0.9601
PKTM	1	74.9844765	74.9844765	12.56	0.0004
Topic Pairs	7	171.6480562	24.5211509	4.11	0.0002
Personal Computer	1	1.3928323	1.3928323	0.23	0.6292
Home Computers	10	55.7242954	5.5724295	0.93	0.5014
Computer Time	12	102.0362612	8.5030218	1.42	0.1483
Computer Expertise	3	53.2202804	17.7400935	2.97	0.0309
State (1:1 Laptop) x FRPL	1	1.1586997	1.1586997	0.19	0.6597
1:1 Laptop x Personal Computer	1	17.4342633	17.4342633	2.92	0.0878
1:1 Laptop x Home Computers	10	60.6331733	6.0633173	1.02	0.4279
1:1 Laptop Computer Time	12	80.8892140	6.7407678	1.13	0.3320
1:1 Laptop x Computer Expertise	3	26.1473483	8,7157828	1.46	0.2240

6.55R-Square	Coefficient of Variation	Root Mean Square Error	Mean Difference of Standardized Scores
0.113636	734.0376	2.443667	0.332908