NISS

Evaluation and Comparison of Two Performance-based Assessments and a Comparable Task-based Assessment of Online Reading and Comprehension

Analysis of ORCA I Data

Weiwei Cui, Eli Bruner, Nell Sedransk

Technical Report 192
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 and Comparison of Two Performance-based Assessments and a Comparable Task-based Assessment of Online Reading and Comprehension

Analysis of ORCA I Data

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

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 I - Executive Summary

ORCA I is the first in a series of large studies to evaluate a performance-based and a comparable task-based assessment of internet literacy. Over 1300 7th graders at 43 schools in two states participated; of these 1079 completed all background information, an offline reading measure and two assessments. Each student was assigned to one of three formats. Two were performance-based assessments defined by their internet access: ORCA-Open with access to the ORCA-Open, ORCA-Closed with access to a synthetic internet built for this project with a search engine and an extensive set of resources. The third format was a task-based multiple choice (MC) assessment that was administered on a computer screen but following a traditional approach; the items themselves mimicked the items in the performance-based assessment.

The overall goal of this study was to investigate the differences between performance-based online assessment and task-based traditional multiple choice assessment of online reading comprehension. The first analyses compared performance-based and task-based assessments (internet, including both ORCA-Open and ORCA-Closed compared to Multiple Choice) and then compared the real-world context of ORCA-Open with the synthetic but stable context of ORCA-Closed. The second set of analyses, reported here, examined the students' performance profiles across the four components (Locate, Evaluate, Synthesize, Communicate) that make up the performance-based assessment. Then the important factors in students' performance on these assessments were identified and subgroups of students who shared common performance profiles where characterized in terms of these factors.

This analysis of data from ORCA I focused on three questions.

1: *Question: What are the overall patterns of performance for each format and what are the dominant factors in performance?*
1a: *Question: For the Communicate component, which factors influence performance?*

2: *Question*: If LESC skills are distinct, how do students' performance profiles differ?

3: *Question: How do students' performance profiles differ? What characterizes subsets of students with similar profiles?*

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

First, it is important to note that the design of ORCA-Open and 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 published *in extenso* elsewhere has confirmed these four components as not fully independent but nonetheless identified with four distinct skills. Thus the

performance-based (ORCA-Open and ORCA-Closed) assessments are multidimensional, that is they draw on multiple distinct skills.

In particular, Locate and Communicate each draw on different skills than Evaluate and Synthesize. This is not evident for the task-based (Multiple Choice) assessment which can be modeled as unidimensional, that is, drawing on a single trait. Further, for the fourth component, Communicate, performance depends on the mode of communication (wiki or email).

For all cases regardless of Format or Version, baseline reading ability (Offline Reading Measure score) dominates as a significant factor and socio-economic status (%FRPL eligible) is also significant.

In addition, Version did matter; in particular one version was dropped from later studies. The importance of Prior Knowledge was evident only on the Multiple Choice assessment and Gender was most important for the internet assessments.

With respect to the individual components of the assessments, Locate was the most distinct task. Two predominant profiles single out the Locate task: either Locate as the most difficult (34-35 % depending on format) or Locate as the easiest (35-38 %). In the context of the Multiple Choice assessment, the profile is even more specific: 47% of students taking the Multiple Choice assessment exhibited one of two patterns: Locate – easiest & Evaluate –most difficult OR Locate – most difficult & Evaluate – easiest. For the internet assessments only 21% followed one of these two patterns.

Introduction: ORCA I – The Students, the Assessments, the Analysis, the Results

Table 1.1a Participated students in 2 States from 43 Schools

		State 1	State 2
State Level	Median Income	$65,573	$46,033
	National Rank	4th	33rd
	1 to 1 Laptop	No	Yes
	Number of schools participating	20	23
	Performance Measure	State classfication	Mean (reading + math scores)
School Level	Performance Level	Every level A-I	Every decile
	Number of 7th graders	68 - 420	12 - 396
	% of free/reduced price lunch	1.7% - 84.1%	10.0% - 77.1%
	%ELL	0% - 12.7%	0% - 25.0%*

Table 1.1b Sample sizes for

		Number of students
State	Laptop State	510
	Non-laptop state	569
Gender	Boys	531
	Girls	548
Format	Multiple Choice	385
	ORCA-Closed	370
	ORCA-Open	324
Total		1079

ORCA I is the first in a series of large studies to evaluate a performance-based and a comparable task-based assessment of internet literacy. Over 1300 7th graders at 43 schools in two states participated; of these 1079 completed all background information, an offline reading measure and two assessments (Table 1.1). Each student was assigned to one of three formats. Two were performance-based assessments defined by their internet access: ORCA-Open with access to the ORCA-Open, ORCA-Closed with access to a synthetic internet built for this project with a search engine and an extensive set of resources. The third format was a task-based multiple choice (MC) assessment that was administered on a computer screen but following a traditional approach with items that mimicked the items in the performance-based assessment.

The overall goal of this study was to investigate the differences between performance-based online assessment and task-based traditional multiple choice assessment of online reading comprehension. One set of key comparisons were between performance-based and task-based assessments (internet, including both ORCA-Open and ORCA-Closed

compared to Multiple Choice) and the comparison of the real-world context of ORCA-Open with the synthetic but stable context of ORCA-Closed. A second set of key analyses examined the factors that affect students' performance, especially the profiles of student's performance across the four components and the characterization of subsets of students who share similar profiles.

Participants in ORCA I

Students came from pairs of 7th grade 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.

Study Design for ORCA I

Eight versions of the assessments (each in all three formats) were developed from scenarios based on science –related topics. Of these, four versions required a final response in an email format; the other four required a final response in the form of a wiki entry. All possible combinations of version pairs that met the constraint of 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 four of the eight versions were assigned (2 wiki response, 2 email response). Within each classroom, each of the possible Version x Format combinations was assigned at random to an approximately equal number of students. Different combinations of versions were assigned to different schools for the first day of testing. The second day the alternate four versions were assigned in each school, again with all possible Version x Format combinations 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 two assessments in that student's single assigned format, with one version requiring a wiki response and the other version requiring one 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), a brief paper and pencil test of relevant prior knowledge (PK) and a survey about personal internet use.

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. Each of the

eight different scenarios (versions) was based on a different science-related research question posed to the student. 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.

The Research Questions

A set of three research questions were posed about the performance-based and the task-based assessments of internet literacy.

1: *Question: What are the overall patterns of performance for each format and what are the dominant factors in performance?*

1a: *Question: For the Communicate component, which factors influence performance?*

2: *Question: If LESC skills are distinct, how do students' performance profiles differ?*

3: *Question: How do students' performance profiles differ? What characterizes subsets of students with similar profiles?*

Analyses responding to these questions make up the remainder of the report; each section is devoted to one question, the response and the supporting analyses.

1: *Question: What are the overall patterns of performance for each format and what are the dominant factors in performance?*

Raw LESC Total Score Distribution
In comparing the Open and ORCA-Closed for the laptop condition, the score differential was significant (8.12 to 7.32); but there was only a minimal, non-significant difference across scores (8.43 to 8.34) for the non-laptop condition. Thus the overall difference between the 1:1 laptop and the non-laptop environments was not attributable to the scenario or the texts but rather to the performance-based nature of the assessment and/or in the accessing of an internet whether actual or synthetic.

The comparison between Multiple Choice and Combined Internet yielded a highly significant score difference both for laptops (11.22 to 7.7) and for non-laptops (11.19 to 8.36).

Comparisons between laptop and non-laptop conditions were mixed (see Table 1.2). For Multiple Choice, the average scores were 11.22 and 11.19, respectively – a non-significant difference. For ORCA-Open, the differences were also non-significant, with laptop students scoring 8.12 and non-laptop students scoring 8.43. The ORCA-Closed displayed a consistent significant difference between the conditions, with the non-laptop students scoring 8.34 and laptop students scoring 7.32. Lastly, for the Combined Internet condition, there was a significant score differential where the laptop students scored an average of 7.7 and the non-laptop students; average score was 8.36.

In sum, there were significant differences between conditions and Internet formats. This was seen especially in the laptop condition and across the Closed and Combined Internets.

Table 1.2

Raw LESC score distribution adjusted for topics

State		Multiple Choice	ORCA-Closed	ORCA-Open	Combined Internet
Non-laptop	Mean	11.19	8.34	8.43	8.36
	Std Dev	1.9	1.72	1.91	1.82
Laptop	Mean	11.22	7.32	8.12	7.7
	Std Dev	1.05	1.55	1.65	1.6

LESC Scores by Gender
Looking at the differences between males and females in their LESC scores, there was no significant difference between genders for the Multiple Choice format. For the ORCA-Closed format, the score differences were significant between males and females both for laptop students (6.32 to 8.32) and for non-laptop students (7.65 to 8.83). There were also significant score differentials for the ORCA-Open format for each laptop condition. For laptop students the male-female a differential was 7.26 to 9.17; and non-laptop students differed with scores of 7.39 to 9.15.

There were no significant scores differences between laptop and non-laptop students when each gender or Internet condition was considered separately as seen in Table 1.3, with the single exception of males in the ORCA-Closed condition (6.32 in the laptop condition and 7.65 in the non-laptop one).

Overall gender differences were significant, especially in the ORCA-Closed condition, generally favoring females.

Table 1.3

Mean LESC scores by Gender adjusted for topics

State	Multiple Choice		ORCA-Closed		ORCA-Open	
	Female	Male	Female	Male	Female	Male
Non-laptop	11.54	10.84	8.83	7.65	9.15	7.39
Laptop	11.36	11.16	8.32	6.32	9.17	7.26

1a: *Question: For the Communicate component, which factors influence performance?*

LESC Scores by Communicate Response: Wiki/Email
The wiki and email modes of communication showed some small differences. All three formats, Multiple Choice, ORCA-Closed, and ORCA-Open for both wiki and email conditions had varied but non-significant score disparities of under 0.43. However, for each mode separately students scored significantly higher on the Multiple Choice (with average scores ranging from 11.41 for Multiple Choice to 7.26 for the Internets).

Some differences between laptop and non-laptop contexts were significant (Table 1.4). First, in the Closed and ORCA-Open formats, non-laptop students performed equivalently or better than did laptop students, but in the Multiple Choice format differences were small and mixed. Also, in the Closed and ORCA-Open condition, wiki scores were significantly higher (8.22 to 7.38 for ORCA-Closed; 8.59 to 7.94). The ORCA-Closed actually had significant differences for the email condition (8.47 to 7.26). The remainder of the comparative differences, Multiple Choice and ORCA-Open email, were non-significant (Tables 1.5).

What is notable here is that the largest differences for any of the subsets of students are between the task-based Multiple Choice assessment and the performance-based Internet assessments. Additionally differences between subsets within a single format did not follow the same pattern for Multiple Choice as for the Internet formats.

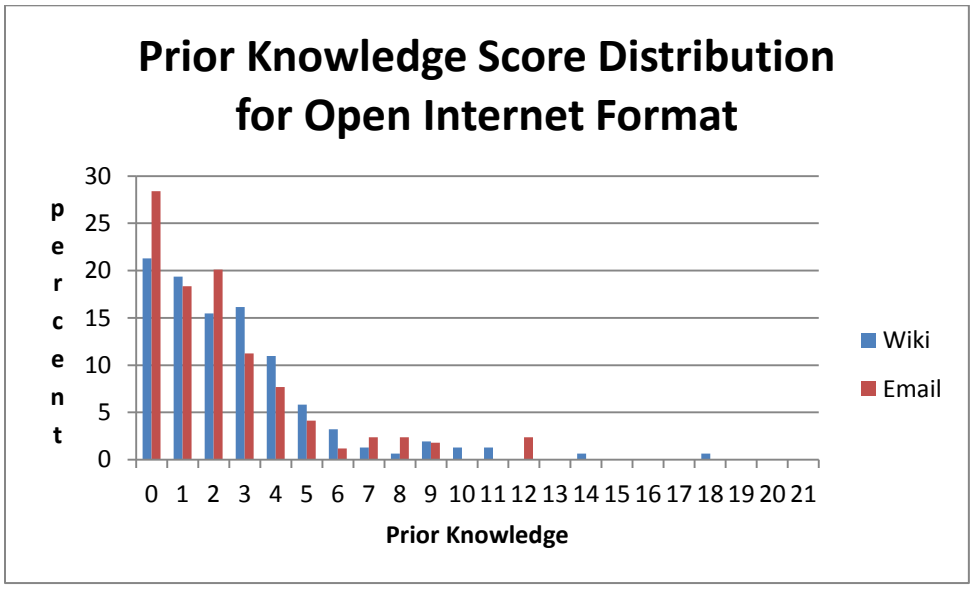
Table 1.4

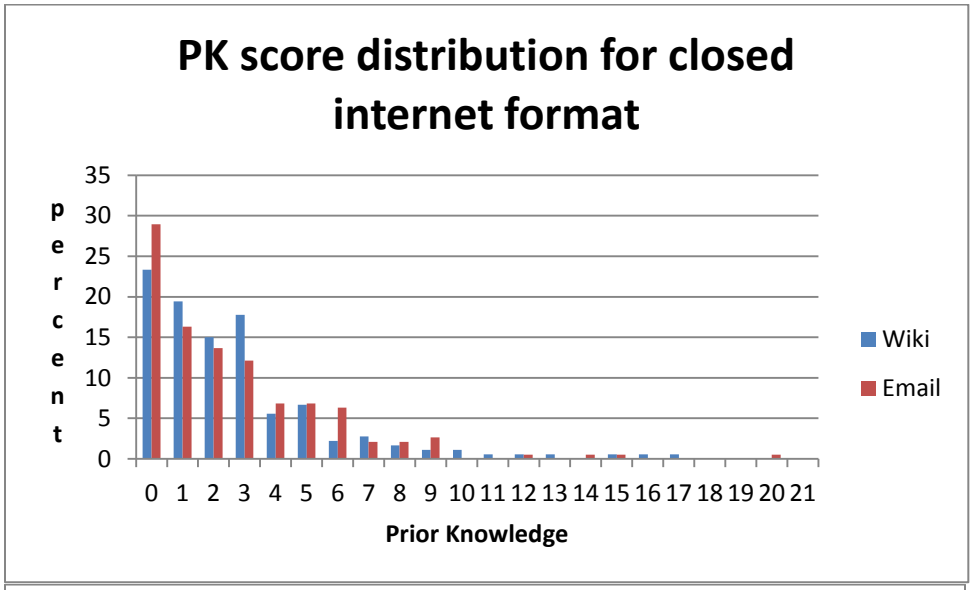
Mean LESC scores by Wiki / Email adjusted for topics

State	Multiple Choice		ORCA-Closed		ORCA-Open	
	Wiki	Email	Wiki	Email	Wiki	Email
Non-laptop	10.99	11.41	8.22	8.47	8.59	8.27
Laptop	11.3	11.14	7.38	7.26	7.94	8.3

Table 1.5

PK	Wiki			Email		
	MC	Closed	Open	MC	Closed	Open
	percent	percent	percent	percent	percent	percent
0	20	23	21	27	29	28
1	10	19	19	11	16	18
2	20	15	15	20	14	20
3	13	18	16	9	12	11
4	10	6	11	9	7	8
5	9	7	6	8	7	4
6	3	2	3	4	6	1
7	3	3	1	3	2	2
8	2	2	1	2	2	2
9	0	1	2	2	3	2
10	1	1	1	1	0	0
11	1	1	1	1	0	0
12	1	1	0	2	1	2
13	1	1	0	0	0	0
14	1	0	1	0	1	0
15	0	1	0	0	1	0
16	1	1	0	0	0	0
17	0	1	0	1	0	0
18	1	0	1	1	0	0
19	1	0	0	0	0	0
20	0	0	0	0	1	0
21	1	0	0	1	0	0
Total Number of Students	200	180	155	183	190	169





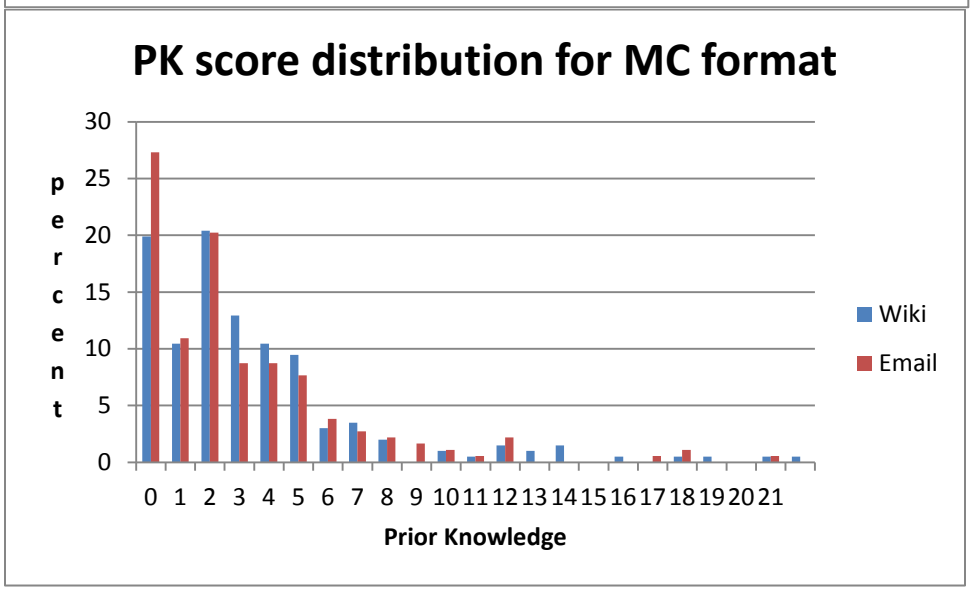
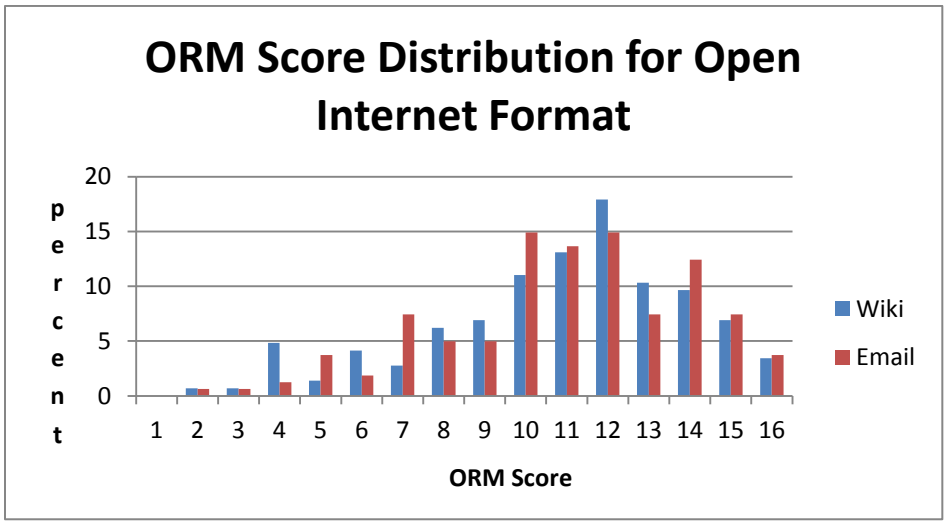
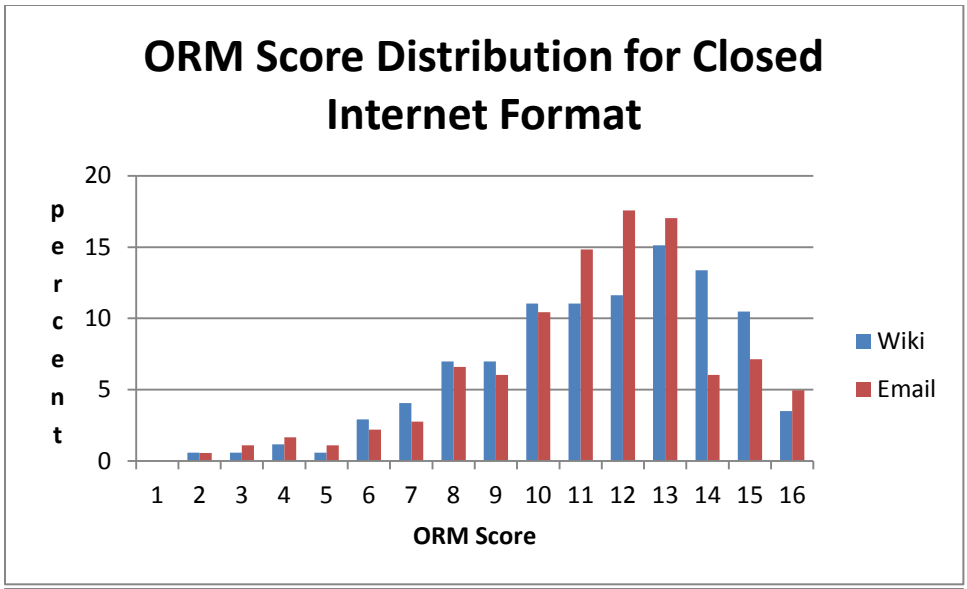


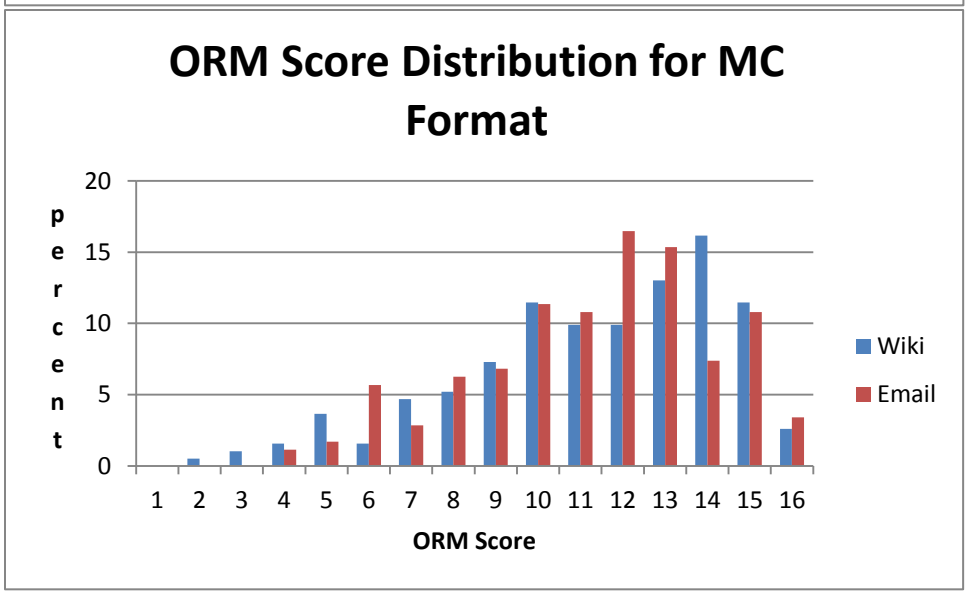
Table 1.6

Offline Reading Measure	Wiki			Email		
	MC	Closed	Open	MC	Closed	Open
	percent	percent	percent	percent	percent	percent
0	0	0	0	0	0	0
1	0	0	0	0	0	0
2	1	1	1	0	1	1
3	1	1	1	0	1	1
4	2	1	5	1	2	1
5	4	1	1	2	1	4
6	2	3	4	6	2	2
7	5	4	3	3	3	7

8	5	7	6	6	7	5
9	7	7	7	7	6	5
10	11	11	11	11	10	15
11	10	11	13	11	15	14
12	10	12	18	16	18	15
13	13	15	10	15	17	7
14	16	13	10	7	6	12
15	11	10	7	11	7	7
16	3	3	3	3	5	4
Total Number of Students	192	172	145	176	182	161







2: *Question*: If LESC skills are distinct, how do students' performance profiles differ? *Which components of the assessment are easiest and which are most difficult?*

Regardless of format, approximately one-third of students found locate to be the easiest (L=4), and one-third found it to be hardest (L=1). For Multiple Choice, ORCA-Closed, and ORCA-Open formats, L = 4 for students was 38%, 38%, and 35%, respectively (Table 2.1).

Looking at Locate scores for MC, fifty percent found either Locate the easiest and Evaluate the hardest, or found Evaluate easiest and Locate to be the hardest (MC = 47%). The same

pairing did not appear as frequently for the internet formats combined with only 21% proportion total for the same pairings.

Table 2.1

Performance Profiles

- IRT scores for L, E, S, C

Ranked for each student (4-digit profile of ranks L,E,S,C)
Highest IRT subscale score (easiest): Rank = 4
Lowest IRT subscale score (most difficult): Rank = 1

- Significant Results:

Locate Easiest or Hardest Component

MC:	38% L=4	and	35% L=1
Internet – Closed:	38% L=4	and	35% L=1
Internet – Open:	35% L=4	and	34% L=1

Most common (4-digit) profiles
4,1, x, x or 1, 4, x, x (L,E or E,L highest & lowest)

MC:	47%
Internet-Combined:	21%

3. *Who are the high-L students (L is easiest)? Who are the low-L students (L is most difficult)?*

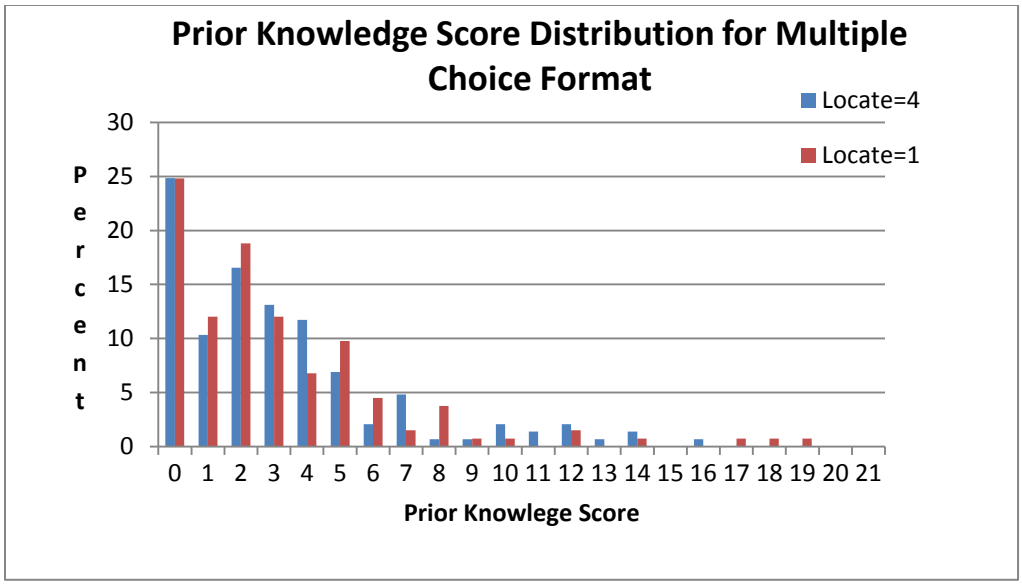
Scoring Profiles - Influence of Prior Knowledge
The most common patterns of student profiles rank Locate as either the easiest (highest subscore) or hardest (lowest subscore) of the four subscores for L, E, S, C. Therefore, these two subsets of students (L=4 and L=1) were analyzed to determine whether clearly identifiable factors drove their performance on the Locate items. In particular, scores on Prior Knowledge were evaluated separately for each of the three formats. . The results was that the score distributions show little difference between the L=4 and the L=1 groups of students for each of the three formats, Multiple Choice, ORCA-Closed and ORCA-Open, (Table 2.2). All three indicated a large proportion of students scoring low (3 points or fewer). This was particularly evident with the ORCA-Open format, which had the highest proportion of low-scoring students as well as a much more truncated tail, i.e., few high scores relative to the other formats.

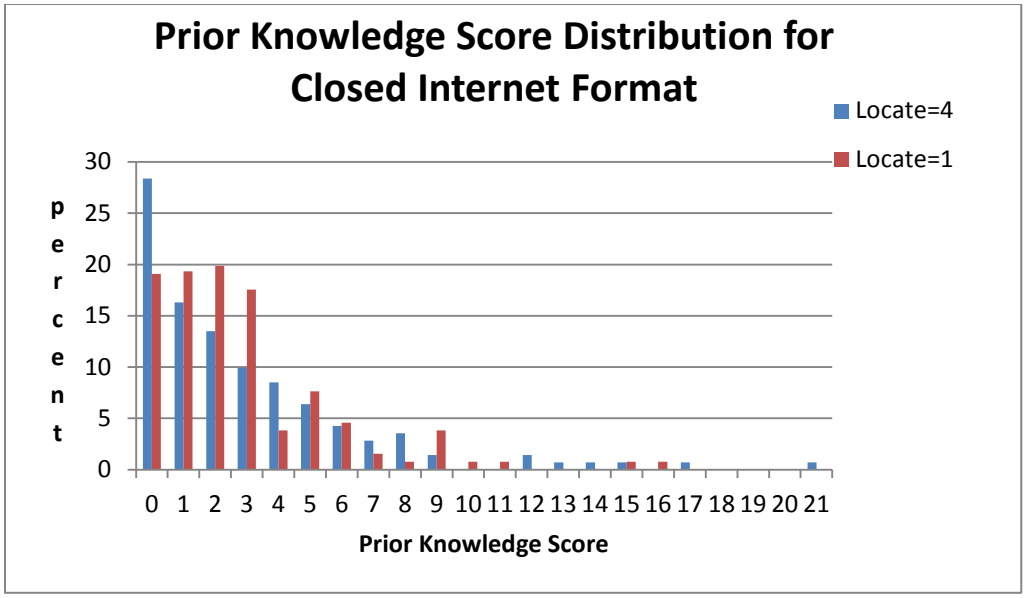
For the Multiple Choice format, the PK score distributions of L =4 and L=1 subgroups were very similar. Both are strongly skewed and bimodal with principal mode at 0 and second mode at PK score=2. Both conditions, oddly enough, had fewer scores of 1 than might have been expected. (Table 2.2) Two conjectures that would offer possible explanations for this are that: 1) there are items, probably a pair of easy items, that are linked in the sense that either both are missed or else both are correctly answered; or 2) the apparent excess of PK scores of zero is due to a distinct subset of students, perhaps poorly prepared or perhaps limited by other factors such as basic reading and language skills or other defining characteristic, who are present within each subgroup.

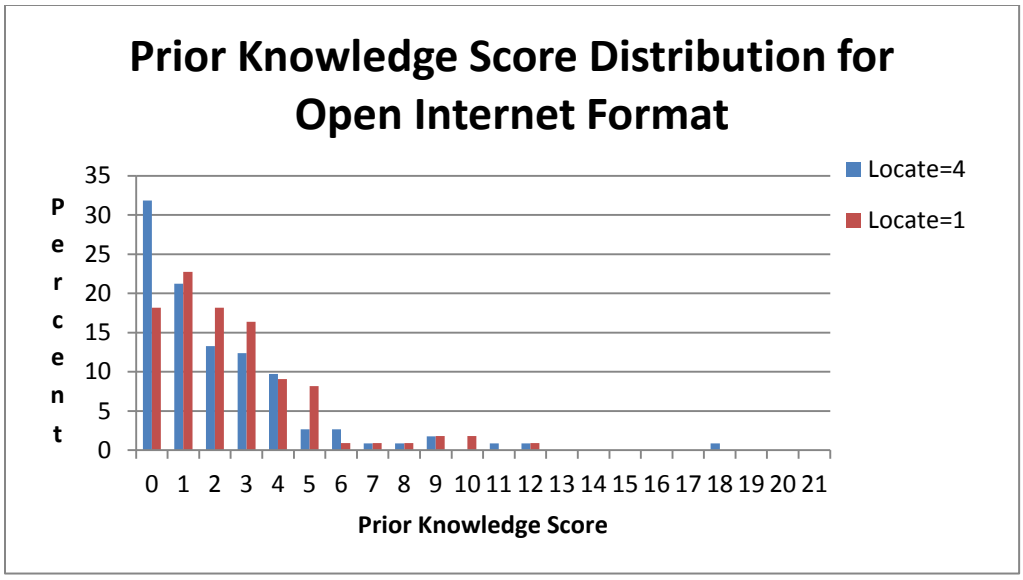
Score distributions for both the ORCA-Closed and the ORCA-Open formats were also strongly skewed, but did not show the same bimodality seen for Multiple Choice scores. Scores for both internet formats were skewed for both subgroups L=1 and L=4, with differences between the two curves in each case attributable to lower average PK score for the L=4 subgroup for both formats (Table 2.2). Overall the proportion of students scoring low (PK score from 0-3) was approximately the same for the two internet formats. So while students with slightly less prior knowledge on average constitute the subgroups with L=4 (Locate as easiest of components), the range of prior knowledge scores is wide for both subgroups L=4 and L=1.

Table 2.2

PK	Rank L=4			Rank L=1		
	MC	Closed	Open	MC	Closed	Open
	percent	percent	percent	percent	percent	percent
0	25	28	32	25	19	18
1	10	16	21	12	19	23
2	17	13	13	19	20	18
3	13	10	12	12	18	16
4	12	9	10	7	4	9
5	7	6	3	10	8	8
6	2	4	3	5	5	1
7	5	3	1	2	2	1
8	1	4	1	4	1	1
9	1	1	2	1	4	2
10	2	0	0	1	1	2
11	1	0	1	0	1	0
12	2	1	1	2	0	1
13	1	1	0	0	0	0
14	1	1	0	1	0	0
15	0	1	0	0	1	0
16	1	0	0	0	1	0
17	0	1	0	1	0	0
18	0	0	1	1	0	0
19	0	0	0	1	0	0
20	0	0	0	0	0	0
21	0	1	0	0	0	0
Total Number of Students	145	141	113	133	131	110


Prior Knowledge Score Distribution for Multiple Choice Format
Locate=4
Locate=1
Percent
30
25
20
15
10
5
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
Prior Knowlege Score


Prior Knowledge Score Distribution for Closed Internet Format
Locate=4
Locate=1
percent
30
25
20
15
10
5
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
Prior Knowledge Score



How are high and low scoring students characterized in terms of offline reading scores?

Scoring Profiles - Influence of Baseline Reading (Offline Reading Measure)
In tracking differences between L=4 and L=1 subgroups across all formats, there were no significant differences noted. There was also almost no significant difference when laptop and non-laptop contexts were considered, with the only significant score differential being for ORCA-Closed for the L=1 subgroup (7.26 to 8.47). Overall however, a significantly larger proportion of 0 ORM scores were found in the L=1 subgroups pooled across all formats.

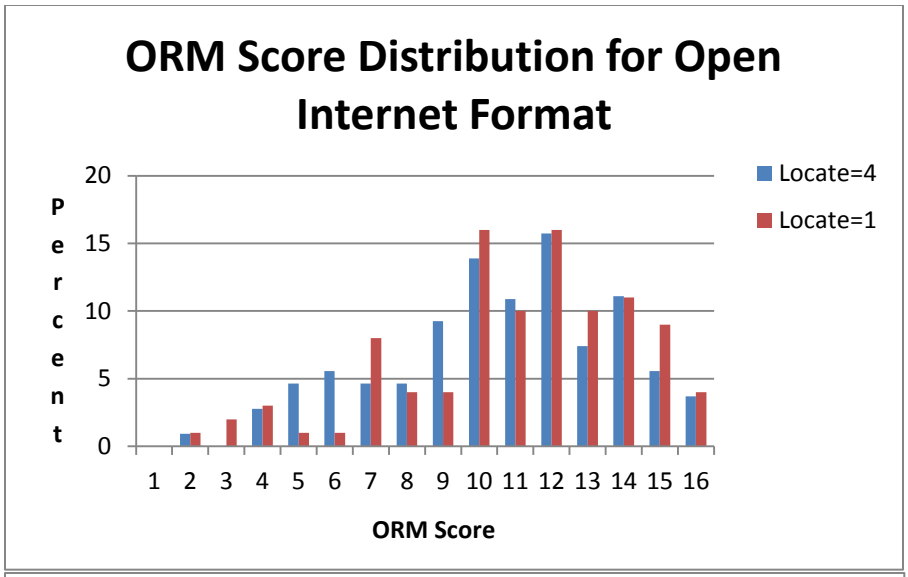
For the Multiple Choice condition, results for both L=4 and L=1 subgroups' ORM scores are approximately normally distributed but truncated at the maximal possible score (Table 2.3). Both the mean and the variance for L=1 are greater for the L=1 subgroup than for L=4 subgroup; the difference in means is significant.

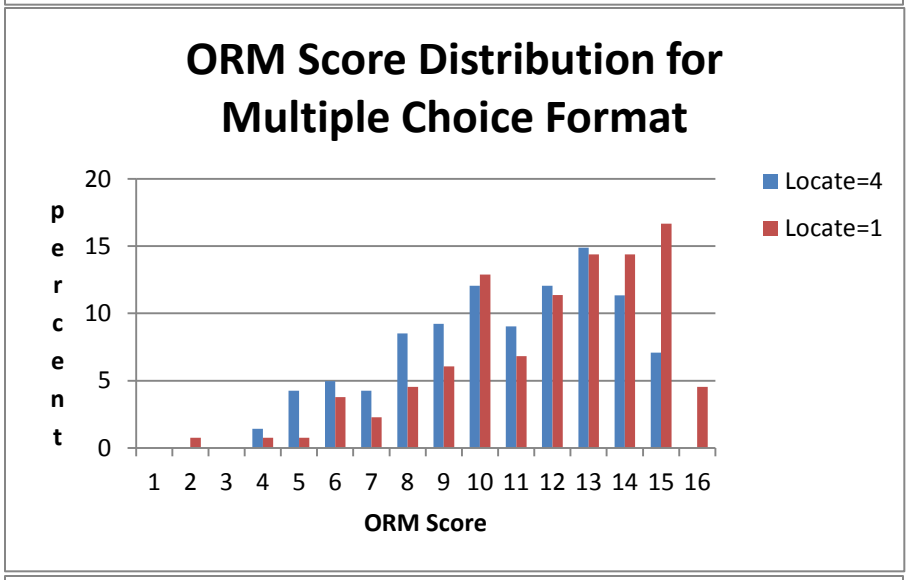
The difference in means was also evident for the ORCA-Closed and ORCA-Open formats, although the magnitude of the differences was smaller and not statistically significant (Tables 2.3). The means for ORM scores for both internet formats were uniformly smaller than those for Multiple Choice; and the variances were slightly larger.

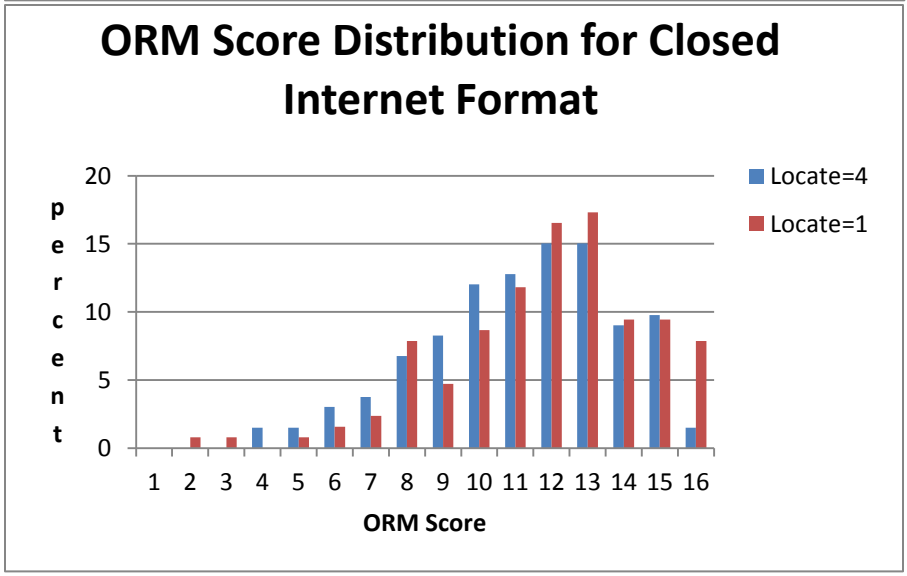
Overall, the persistent difference was the higher mean ORM scores for the L-1 subgroup over the L=4 subgroup, although the magnitude was small for the internet formats.

Table 2.3

Offline Reading Measure	Rank L=4			Rank L=1		
	MC	Closed	Open	MC	Closed	Open
	percent	percent	percent	percent	percent	percent
0	0	0	0	0	0	0
1	0	0	0	0	0	0
2	0	0	1	1	1	1
3	0	0	0	0	1	2
4	1	2	3	1	0	3
5	4	2	5	1	1	1
6	5	3	6	4	2	1
7	4	4	5	2	2	8
8	9	7	5	5	8	4
9	9	8	9	6	5	4
10	12	12	14	13	9	16
11	9	13	11	7	12	10
12	12	15	16	11	17	16
13	15	15	7	14	17	10
14	11	9	11	14	9	11
15	7	10	6	17	9	9
16	0	2	4	5	8	4
Total Number of Students	141	133	108	132	127	100


ORM Score Distribution for Open Internet Format
Percent
20
15
10
5
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
ORM Score
Locate=4
Locate=1


ORM Score Distribution for Multiple Choice Format
percent
20
15
10
5
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
ORM Score
Locate=4
Locate=1


ORM Score Distribution for Closed Internet Format
percent
20
15
10
5
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
ORM Score
Locate=4
Locate=1

3: *Question: How do students' performance profiles differ? What characterizes subsets of students with similar profiles?*

When the factors driving students' performances are the same for almost all students, general models can work well to describe these. However, if the relative importance of factors differs among students, general models will average across these so that they describe no subgroup of students very well. Partitioning algorithms such as regression tree analysis is useful to define subsets of students who perform similarly, as well as to determine the factors that affect subsets' performances. Classification And Regression Trees (CARTs) are specifically useful in that they allow for the creation of subsets within subsets based on score that are defined by different cutpoints for key factors or different sets of factors. Thus, this sequential subsetting process can refocus further and further until groups' minimum homogeneous sizes have been established.

Recursive Partitioning Procedure

Regression Tree analysis is a stepwise partitioning algorithm. This analysis proceeds based on regression by determining the single most important factor to define a cutpoint that splits the group (or later on a subgroup) into two smaller subgroups. This split maximizes the within subgroup homogeneity and between subgroup differences (i.e., minimizes within subgroup sum or squares, maximizing reduction in variance due to clustering). The presentation of the splits in the figure shows the lower values of the splitting factor to the left, the values above the cutpoint to the right.

At the next step each of the subgroups formed is treated individually, once again by determining the single most important factor to define the best cut point to split that subgroup into two smaller parts. A critical feature of this process is that different factors may or may not be used to split different subgroups at the same stage because splitting is independent across subgroups. This permits differential dependence of subgroups on specific sets of factors and thus accommodates interactions among factors in a natural way.

The tree terminates in its final subgroups when further splitting fails to reduce variance or when the subgroup sizes are too small for reliable partitioning.

The results for one version, Second Hand Smoke, are analyzed using CART for each of the formats separately.

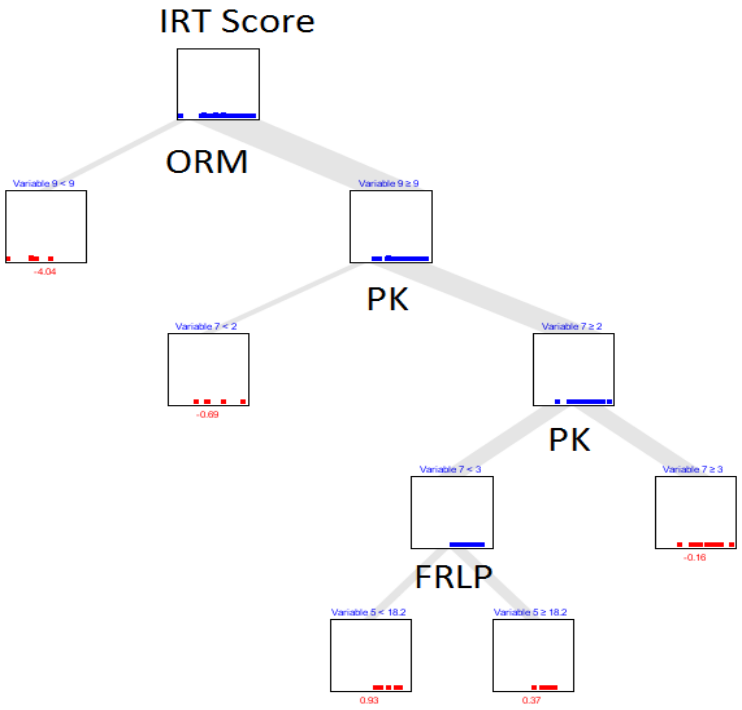
Regression Tree for Multiple Choice

The Multiple Choice CART subdivided students' IRT scores, subset-by-subset as shown in (Table 3.1). The initial division is based on Offline Reading Measurement (ORM), followed by prior knowledge (PK). These first two splits separate the very weak students from the rest, making two sufficiently distinctive subgroups that these do not split further. The next division of the larger group of students again divides based on prior knowledge. This is followed by further subdivision of the mid-range and higher scores based on socioeconomic status (FRPL). What this indicates is that prior knowledge is the primary strong factor for all except the lowest IRT scores, but that for equivalent reading skills prior knowledge is next in importance. Socioeconomic status has a lower-level impact for

students with mid-range reading skills. Ergo, while basic reading is the single best indicator for multiple-choice performance-based on IRT score, some subsets of students can only be described when both prior knowledge and socioeconomic status are taken into account.

Table 3.1

Regression Tree (second-hand smoke) Multiple Choice Format

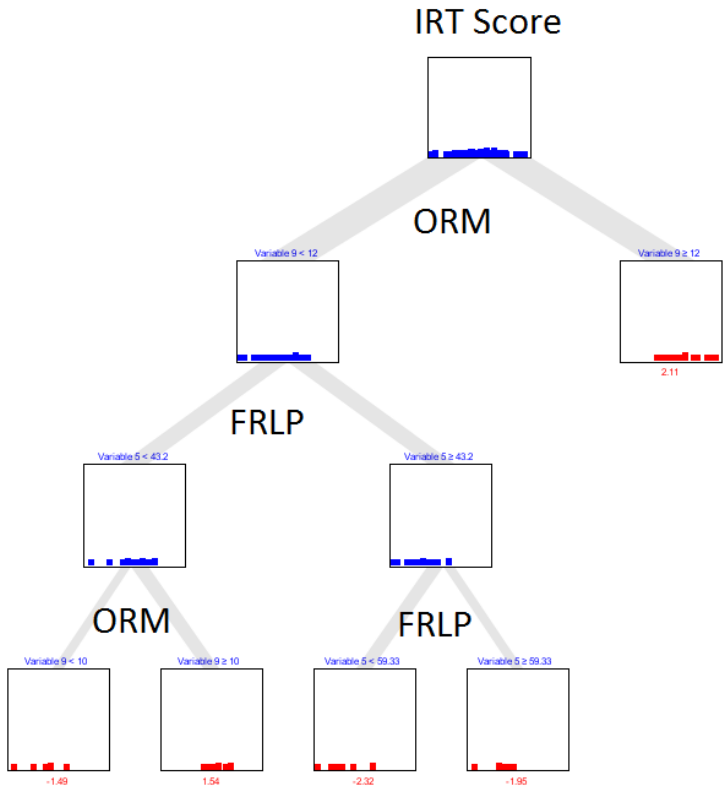


Regression Tree for ORCA-Closed

Like the Multiple Choice CART, the first division CART made for ORCA-Closed was ORM. This separated a high-performing subgroup that did not subdivide further. The majority were subdivided into three groups according to socioeconomics (%FRPL eligible) of the schools. The most privileged schools were subdivided again according to baseline reading, ORM scores (Table 3.2). Thus, apart from the highest performing students, socioeconomics is the most important factor with respect to IRT scores.

Table 3.2

Regression Tree (second-hand smoke) ORCA-Closed Format



Regression Tree for ORCA-Open

The ORCA-Open CART subdivides initially by socioeconomic indicator, %FRPL eligible. (Table 3.3). The decreasing subdivision from ORM scores was into PK, subdivided twice. For the more privileged schools (FRPL eligible < 40%), prior knowledge is the factor with the next highest impact. In fact, division according to %FRPL creates three final subgroups. For the less privileged schools, gender is the factor of next importance after %FRPL, and the only one required to define the final subgroups.

Table 3.3

Regression Tree (second-hand smoke) for ORCA-Open Format