Matthew Barbee
SLS: Testing 490
Instructor: Dr. Hudson
February 22, 2013

Language Test Evaluation

I. Item Analysis

Forty-five students (*n*=45) were given both a listening test and a reading test. Both tests had 30 questions each (*k*=30). Each question on the reading test had four distractors. The following tables and figures represent the item analysis and descriptive analysis of both tests.

I will begin my analysis by examining the data from the reading test. Because only one administration of the test was given and because a pass/fail cutoff point was not established, it is difficult to analyze the test as if it is a criterion referenced test, or CRT. For that reason, most of my analysis will treat both tests as if each is a norm-referenced test, or NRT. That said, when analyzing NRTs, item facility (IF) and item discrimination (ID) are important points of reference. For this analysis, I used the following scales for IF and ID to help determine if items were well written, needed some change, or should be rejected. The scales I used for IF and ID are as follows:

Item Facility	
0.7 – 1.0	Low difficulty
0.3 – 0.7	Normal difficulty
0 – 0.3	High difficulty

Item Discrimination	
0.4 – 1.0	Very good discrimination
0.3 – 0.39	Good discrimination
0.2 – 0.29	Marginal discrimination
0 – 0.19	Minimal discrimination (reject)

Table 1 in the Appendix shows the item analysis for the reading test. First, looking at the IF, we can see that items 3, 5, 6, 9, 10, 12, 17, 19, 20, 21, 23, 24, 25, and 27 have low difficulty, (items 18 and 23 with IFs of 0.71 and 0.73 respectively have been left out of this list at my discretion). However, a high IF alone is not enough to disregard or change an item. We must also look at each items' ID. While a majority of the items fall into the "very good" and "good" categories (60%), 6 items (20%) show marginal discrimination, and 6 items (20%) show minimal discrimination. A breakdown of the items that provide marginal and minimal discrimination is as follows:

Marginal	Minimal
3	6
12	9
17	10
18	21
23	24
28	27

The information provided by both the IF and the ID allows me to make preliminary recommendations for items to be divided into three categories, (a) keep, (b) examine further and possibly change, and (c) reject. Items that are flagged under both the IF and ID analysis fall under the rejected category; items that are flagged as being deficient in either IF or ID both not in both may need some changes; and items that show normal IF and ID should be kept. Those preliminary recommendations are as follows:

Keep	Change	Reject
1, 2, 4, 7, 8, 11, 13, 14, 15, 16, 22, 26, 29, 30	5, 18, 19, 20, 25, 28	3, 6, 9, 10, 12, 17, 21, 23, 24, 27

At first glance, it does seem that there are a lot of items that I am recommending be rejected (30%). This may be due to the fact that I am examining the data from the reading test through a NRT lens, when in fact the results may be more in alignment with a CRT.

When looking at the distribution of the scores in Figure 2 of the Appendix, we can see that the distribution looks more like results from a CRT post-test because of its negative distribution. If the reading test is intended to be a NRT, the items above that fall in the "change" and "rejected" categories would have to be examined much more closely. Whether they are changed or rejected, these items would have to provide more discrimination and higher difficulty to better fulfill the purpose of a NRT.

To provide a closer examination, we can turn to a distractor analysis of each item (see Table 2 in the Appendix). For this analysis I used the spreadsheet application Excel and calculated the percentage of how each ability group (high and low) answered each alternative distractor. For the distractor analysis, I focused on individual distractors that provided no distraction, those that where chosen 0% of the time by both the high and low groups. Those distractors are 3b, 5a, 12a and b, 16c, 17d, 20a, 22a, 23a, 24a and b, 27a, and 30b. When comparing these distractors to the preliminary "change" and "reject" items, we can see that all of them overlap with the exception of 16c, 22a, and 30b. This means that items 5 and 20 can safely remain in the "change" category, while 3, 12, 17, 23, 24, and 27 can be moved to the "change" category, knowing that individual distractors for each of these items must be altered. It should be noted here that while items 16, 22, and 30 also had distractors that provided 0% distraction, they didn't present a problem in regards to IF or ID.

Looking at another issue in regards to the distractor analysis, some alternative distractors caused more students from the higher scoring group rather than the lower scoring group to choose them. The analysis provided by the TAP program for the same data set was used to find this information (see Table 3 in the Appendix). Those distractors are 1c, 10a, 11b, 13c, 16c, 18d, and 21a. Of these, it is already recommended that item 18

be changed and that items 10 and 21 be rejected. While the other distractors don't cause a problem to IF and ID, some attention may be needed to better distribute scores in the case of a NRT. Also, if this were instead a CRT, an instructor may want to go back and check the content of the course with how each distractor is working within these items.

Overall, the final recommendations for the changes to the reading test items are as follows:

Keep	Change	Reject
1, 2, 4, 7, 8, 11, 13, 14, 15, 16, 22, 26, 29, 30	3*, 5*, 12*, 17*, 18*, 19, 20*, 23*, 24*, 25, 27*, 28	6, 9, 10, 21

* Those items that have distractors that need to be changed.

II. Descriptive Statistics

The following table represents the descriptive statistics from the listening and reading subtests and overall test. Following the table are the histograms for each subtest and the total test (Figures 1, 2, and 3). A short analysis of the distribution patterns will follow each graph.

	N	*k*	*M*	Mode	Median	Midpoint	SD	Var.	Low	High	Range
Listening	45	30	**17.71**	19.00	18.00	16.00	**4.26**	18.16	7.00	25.00	19.00
Reading	45	30	**22.20**	24.00	23.00	19.00	**4.79**	22.96	8.00	30.00	23.00
Total	45	60	**39.91**	43.00	41.00	36.00	**8.45**	71.46	19.00	53.00	35.00

At first glance, we can clearly see central tendency in both subtests and on the combined scores from both subtests—mean, mode, median, and mode being nearly the same. However, there is still some abnormal dispersion as evidenced by the standard deviation of scores from the mean as well as the range for each test. When we graph all three, more information can be gleaned.

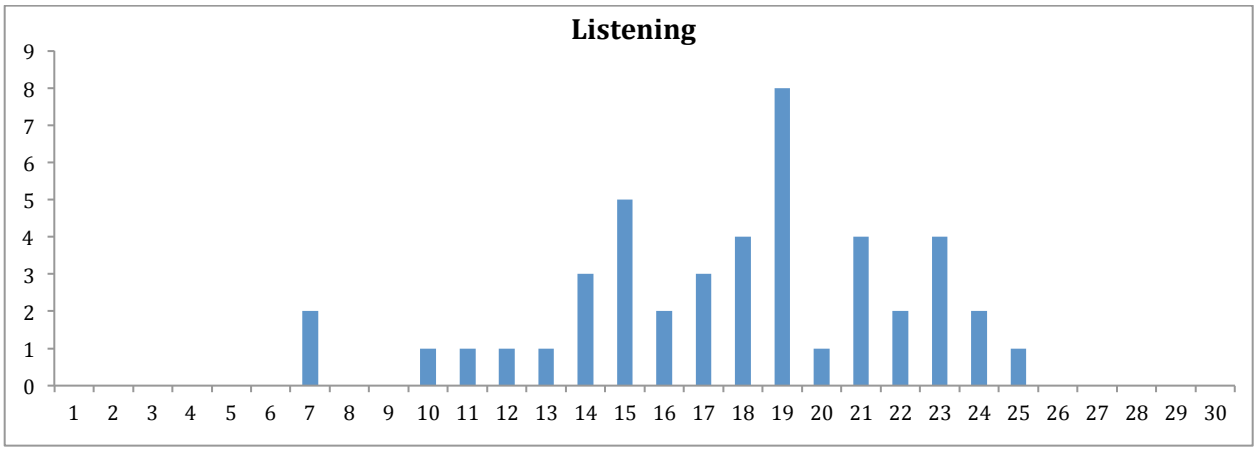

Figure 1. Histogram of the listening test.

We can see from the histogram of the listening test (Figure 1) that while there is a slight negative distribution, the is a general normal distribution to the scores. This pattern may be indicative of a NRT. There is a strong central tendency with almost even distribution on both sides of the mean. There is a little concern that the lowest scores fall just beyond two standard deviations away from the mean while the upper scores are within two standard deviations. Consideration could be taken to achieve a more normalized distribution.

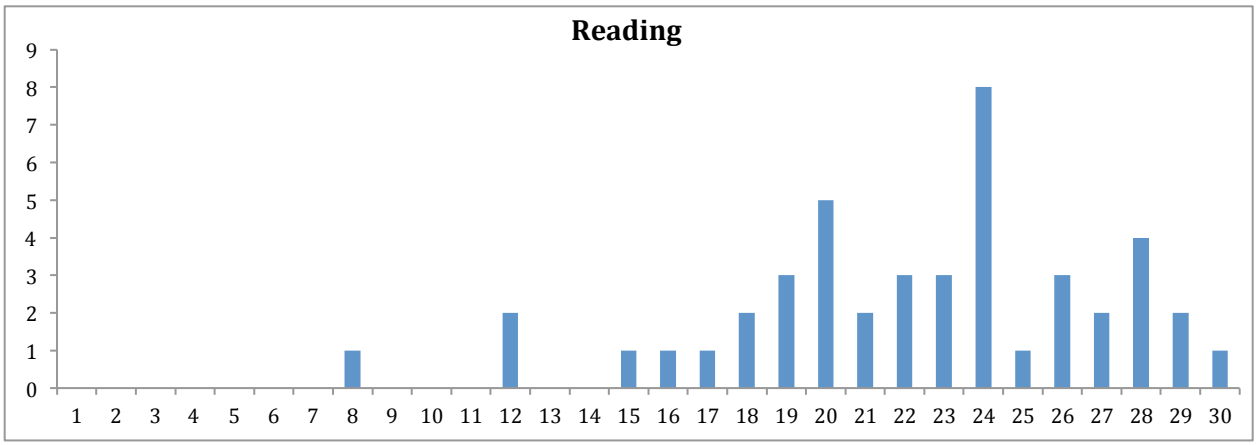

Figure 2. Histogram of the reading test.

The reading subtest (Figure 2) shows even more of a negative distribution than the listening test and somewhat resembles what a teacher may look for in the results of a CRT post-test. Here we can see that the lowest scores fall approximately three standard

deviations away from the mean while the upper scores are just beyond only one standard deviation. If this was a CRT post-test, a teacher should be very concerned for the low achieving students as they would not have met the criterion set for the class.

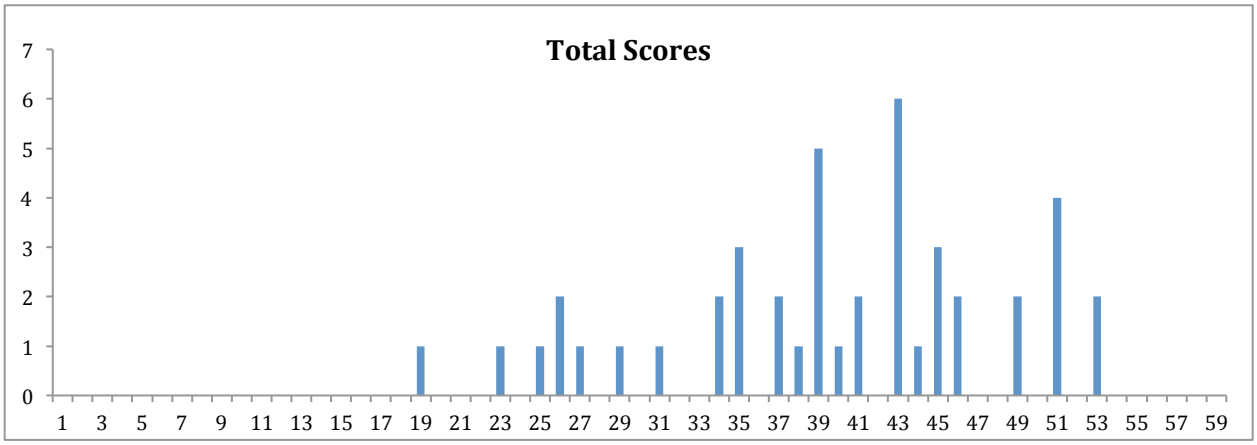

Figure 3. Histogram of the total scores from the listening test and the reading test.

Because the histogram of the total scores for the both tests (Figure 3) is, in a sense, an combination of the statistics described in the above subtests, it is clear that the findings will fall somewhere between what was concluded for the subtests as well. That said, while there is a negative distribution on the total scores, a central tendency can still be seen.

Overall, when scores from both subtests are combined, ambiguity remains as to whether the test was meant as a NRT or a CRT. Claims could be made for both. If the intention was to create a NRT, questions would have to be revised/rejected and the overall difficulty of the test would have to increase slightly.

Appendix

Table 1

Item Analysis

Item	Item Facility	Item Facility Upper	Item Facility Lower	Item Discrimination
1	0.69	0.87	0.53	0.33
2	0.64	0.87	0.27	0.60
3	**0.91**	**1.00**	**0.73**	**0.27**
4	0.53	0.87	0.20	0.67
5	0.89	1.00	0.67	0.33
6	**0.84**	**0.87**	**0.87**	**0.00**
7	0.58	0.87	0.27	0.60
8	0.53	0.93	0.40	0.53
9	**0.89**	**0.93**	**0.80**	**0.13**
10	**0.84**	**0.80**	**0.73**	**0.07**
11	0.62	0.80	0.47	0.33
12	**0.93**	**1.00**	**0.80**	**0.20**
13	0.47	0.67	0.20	0.47
14	0.69	1.00	0.40	0.60
15	0.60	0.87	0.47	0.40
16	0.76	0.93	0.60	0.33
17	**0.89**	**0.93**	**0.73**	**0.20**
18	0.71	0.87	0.60	0.27
19	0.80	1.00	0.60	0.40
20	0.84	1.00	0.60	0.40
21	**0.89**	**0.87**	**0.80**	**0.07**
22	0.51	0.80	0.40	0.40
23	0.87	0.93	0.73	0.20
24	**0.96**	**1.00**	**0.87**	**0.13**
25	0.82	1.00	0.67	0.33
26	0.73	0.93	0.47	0.47
27	**0.82**	**0.80**	**0.73**	**0.07**
28	0.67	0.80	0.53	0.27
29	0.67	0.93	0.40	0.53
30	0.60	0.87	0.33	0.53

Appendix

Table 2

Distractor Analysis

Item	Answer Key	Group	A	B	C	D
1	a	high	0.87	0.07	0.07	0.00
		low	0.53	0.13	0.13	0.20
2	c	high	0.00	0.13	0.87	0.00
		low	0.07	0.33	0.27	0.27
3	d	high	0.00	0.00	0.00	1.00
		low	0.07	0.00	0.20	0.60
4	c	high	0.00	0.07	0.87	0.07
		low	0.27	0.13	0.20	0.27
5	b	high	0.00	1.00	0.00	0.00
		low	0.00	0.60	0.07	0.20
6	d	high	0.07	0.07	0.00	0.87
		low	0.07	0.00	0.07	0.67
7	a	high	0.87	0.00	0.07	0.13
		low	0.27	0.07	0.07	0.47
8	a	high	0.93	0.00	0.07	0.13
		low	0.40	0.07	0.13	0.33
9	d	high	0.07	0.00	0.00	0.93
		low	0.07	0.07	0.07	0.60
10	c	high	0.07	0.07	0.80	0.07
		low	0.00	0.07	0.60	0.20
11	a	high	0.80	0.07	0.07	0.07
		low	0.47	0.00	0.27	0.27
12	d	high	0.00	0.00	0.00	1.00
		low	0.00	0.00	0.20	0.67
13	b	high	0.07	0.60	0.20	0.07
		low	0.27	0.20	0.13	0.27
14	c	high	0.00	0.07	0.87	0.07
		low	0.20	0.20	0.33	0.20
15	a	high	0.87	0.00	0.07	0.13
		low	0.47	0.07	0.13	0.27
16	b	high	0.00	0.93	0.00	0.13
		low	0.07	0.53	0.00	0.33
17	c	high	0.00	0.07	0.93	0.00
		low	0.07	0.13	0.67	0.00
18	b	high	0.00	0.80	0.00	0.20
		low	0.13	0.53	0.07	0.13
19	b	high	0.00	1.00	0.00	0.07
		low	0.20	0.60	0.07	0.07
20	b	high	0.00	1.00	0.00	0.00
		low	0.00	0.53	0.27	0.13
21	c	high	0.07	0.00	0.87	0.07
		low	0.00	0.07	0.67	0.13
22	c	high	0.00	0.07	0.73	0.13
		low	0.00	0.13	0.33	0.40
23	b	high	0.00	0.93	0.00	0.07
		low	0.00	0.67	0.07	0.20
24	d	high	0.00	0.00	0.00	1.00
		low	0.00	0.00	0.07	0.73
25	c	high	0.00	0.00	1.00	0.00
		low	0.07	0.07	0.53	0.20
26	a	high	0.93	0.07	0.00	0.00
		low	0.47	0.07	0.13	0.27
27	c	high	0.00	0.07	0.80	0.13
		low	0.00	0.13	0.60	0.13
28	b	high	0.00	0.80	0.07	0.13
		low	0.07	0.47	0.20	0.20
29	c	high	0.07	0.00	0.93	0.07
		low	0.07	0.33	0.40	0.13
30	a	high	0.87	0.00	0.00	0.07
		low	0.33	0.00	0.33	0.13

Appendix

Table 3
TAP Distractor Analysis

Item	Group	Option 1	Option 2	Option 3	Option 4
1	TOTAL	31*(0.689)	3 (0.067)	6 (0.133)	5 (0.111)
	High	17 (0.810)	1 (0.048)	3 (0.143)	0 (0.000)
	Low	9 (0.563)	2 (0.125)	2 (0.125)	3 (0.188)
	Diff	8 (0.247)	-1(-0.077)	1 (0.018)	-3(-0.188)
2	TOTAL	1 (0.022)	9 (0.200)	29*(0.644)	6 (0.133)
	High	0 (0.000)	2 (0.095)	19 (0.905)	0 (0.000)
	Low	1 (0.063)	5 (0.313)	5 (0.313)	5 (0.313)
	Diff	-1(-0.063)	-3(-0.217)	14 (0.592)	-5(-0.313)
3	TOTAL	1 (0.022)	0 (0.000)	3 (0.067)	41*(0.911)
	High	0 (0.000)	0 (0.000)	0 (0.000)	21 (1.000)
	Low	1 (0.063)	0 (0.000)	3 (0.188)	12 (0.750)
	Diff	-1(-0.063)	0 (0.000)	-3(-0.188)	9 (0.250)
4	TOTAL	7 (0.156)	7 (0.156)	24*(0.533)	7 (0.156)
	High	0 (0.000)	2 (0.095)	18 (0.857)	1 (0.048)
	Low	4 (0.250)	4 (0.250)	3 (0.188)	5 (0.313)
	Diff	-4(-0.250)	-2(-0.155)	15 (0.670)	-4(-0.265)
5	TOTAL	0 (0.000)	40*(0.889)	1 (0.022)	4 (0.089)
	High	0 (0.000)	21 (1.000)	0 (0.000)	0 (0.000)
	Low	0 (0.000)	11 (0.688)	1 (0.063)	4 (0.250)
	Diff	0 (0.000)	10 (0.313)	-1(-0.063)	-4(-0.250)
6	TOTAL	3 (0.067)	3 (0.067)	1 (0.022)	38*(0.844)
	High	1 (0.048)	1 (0.048)	0 (0.000)	19 (0.905)
	Low	1 (0.063)	0 (0.000)	1 (0.063)	14 (0.875)
	Diff	0(-0.015)	1#(0.048)	-1(-0.063)	5 (0.030)
7	TOTAL	26*(0.578)	2 (0.044)	3 (0.067)	14 (0.311)
	High	16 (0.762)	1 (0.048)	1 (0.048)	3 (0.143)
	Low	4 (0.250)	1 (0.063)	2 (0.125)	9 (0.563)
	Diff	12 (0.512)	0(-0.015)	-1(-0.077)	-6(-0.420)
8	TOTAL	24*(0.533)	3 (0.067)	7 (0.156)	11 (0.244)
	High	16 (0.762)	0 (0.000)	2 (0.095)	3 (0.143)
	Low	6 (0.375)	1 (0.063)	3 (0.188)	6 (0.375)
	Diff	10 (0.387)	-1(-0.063)	-1(-0.092)	-3(-0.232)
9	TOTAL	2 (0.044)	1 (0.022)	2 (0.044)	40*(0.889)
	High	1 (0.048)	0 (0.000)	1 (0.048)	19 (0.905)
	Low	1 (0.063)	1 (0.063)	1 (0.063)	13 (0.813)
	Diff	0(-0.015)	-1(-0.063)	0(-0.015)	6 (0.092)
10	TOTAL	1 (0.022)	2 (0.044)	38*(0.844)	4 (0.089)
	High	1 (0.048)	1 (0.048)	18 (0.857)	1 (0.048)
	Low	0 (0.000)	1 (0.063)	12 (0.750)	3 (0.188)
	Diff	1 (0.048)	0(-0.015)	6 (0.107)	-2(-0.140)
11	TOTAL	28*(0.622)	3 (0.067)	8 (0.178)	6 (0.133)
	High	16 (0.762)	2 (0.095)	2 (0.095)	1 (0.048)
	Low	8 (0.500)	0 (0.000)	4 (0.250)	4 (0.250)
	Diff	8 (0.262)	2 (0.095)	-2(-0.155)	-3(-0.202)
12	TOTAL	0 (0.000)	0 (0.000)	3 (0.067)	42*(0.933)
	High	0 (0.000)	0 (0.000)	0 (0.000)	21 (1.000)
	Low	0 (0.000)	0 (0.000)	3 (0.188)	13 (0.813)
	Diff	0 (0.000)	0 (0.000)	-3(-0.188)	8 (0.188)
13	TOTAL	6 (0.133)	21*(0.467)	7 (0.156)	11 (0.244)
	High	2 (0.095)	14 (0.667)	3 (0.143)	2 (0.095)
	Low	4 (0.250)	4 (0.250)	2 (0.125)	6 (0.375)
	Diff	-2(-0.155)	10 (0.417)	1 (0.018)	-4(-0.280)
14	TOTAL	3 (0.067)	5 (0.111)	31*(0.689)	6 (0.133)
	High	0 (0.000)	1 (0.048)	19 (0.905)	1 (0.048)
	Low	3 (0.188)	3 (0.188)	7 (0.438)	3 (0.188)
	Diff	-3(-0.188)	-2(-0.140)	12 (0.467)	-2(-0.140)
15	TOTAL	27*(0.600)	3 (0.067)	5 (0.111)	10 (0.222)
	High	16 (0.762)	0 (0.000)	2 (0.095)	3 (0.143)
	Low	7 (0.438)	1 (0.063)	2 (0.125)	6 (0.375)
	Diff	9 (0.324)	-1(-0.063)	0(-0.030)	-3(-0.232)

Item	Group	Option 1	Option 2	Option 3	Option 4
16	TOTAL	1 (0.022)	34*(0.756)	2 (0.044)	8 (0.178)
	High	0 (0.000)	17 (0.810)	1 (0.048)	3 (0.143)
	Low	1 (0.063)	10 (0.625)	0 (0.000)	5 (0.313)
	Diff	-1(-0.063)	7 (0.185)	1 (0.048)	-2(-0.170)
17	TOTAL	1 (0.022)	4 (0.089)	40*(0.889)	0 (0.000)
	High	0 (0.000)	1 (0.048)	20 (0.952)	0 (0.000)
	Low	1 (0.063)	3 (0.188)	12 (0.750)	0 (0.000)
	Diff	-1(-0.063)	-2(-0.140)	8 (0.202)	0 (0.000)
18	TOTAL	2 (0.044)	32*(0.711)	3 (0.067)	8 (0.178)
	High	0 (0.000)	17 (0.810)	0 (0.000)	4 (0.190)
	Low	2 (0.125)	9 (0.563)	3 (0.188)	2 (0.125)
	Diff	-2(-0.125)	8 (0.247)	-3(-0.188)	2 (0.065)
19	TOTAL	3 (0.067)	36*(0.800)	4 (0.089)	2 (0.044)
	High	0 (0.000)	20 (0.952)	0 (0.000)	1 (0.048)
	Low	3 (0.188)	10 (0.625)	2 (0.125)	1 (0.063)
	Diff	-3(-0.188)	10 (0.327)	-2(-0.125)	0(-0.015)
20	TOTAL	1 (0.022)	38*(0.844)	4 (0.089)	2 (0.044)
	High	0 (0.000)	21 (1.000)	0 (0.000)	0 (0.000)
	Low	0 (0.000)	10 (0.625)	4 (0.250)	2 (0.125)
	Diff	0 (0.000)	11 (0.375)	-4(-0.250)	-2(-0.125)
21	TOTAL	1 (0.022)	1 (0.022)	40*(0.889)	3 (0.067)
	High	1 (0.048)	0 (0.000)	19 (0.905)	1 (0.048)
	Low	0 (0.000)	1 (0.063)	13 (0.813)	2 (0.125)
	Diff	1 (0.048)	-1(-0.063)	6 (0.092)	-1(-0.077)
22	TOTAL	2 (0.044)	6 (0.133)	23*(0.511)	14 (0.311)
	High	2 (0.095)	2 (0.095)	14 (0.667)	3 (0.143)
	Low	0 (0.000)	3 (0.188)	6 (0.375)	7 (0.438)
	Diff	2 (0.095)	-1(-0.092)	8 (0.292)	-4(-0.295)
23	TOTAL	1 (0.022)	39*(0.867)	1 (0.022)	4 (0.089)
	High	0 (0.000)	20 (0.952)	0 (0.000)	1 (0.048)
	Low	0 (0.000)	12 (0.750)	1 (0.063)	3 (0.188)
	Diff	0 (0.000)	8 (0.202)	-1(-0.063)	-2(-0.140)
24	TOTAL	0 (0.000)	0 (0.000)	2 (0.044)	43*(0.956)
	High	0 (0.000)	0 (0.000)	0 (0.000)	21 (1.000)
	Low	0 (0.000)	0 (0.000)	2 (0.125)	14 (0.875)
	Diff	0 (0.000)	0 (0.000)	-2(-0.125)	7 (0.125)
25	TOTAL	1 (0.022)	1 (0.022)	37*(0.822)	6 (0.133)
	High	0 (0.000)	0 (0.000)	20 (0.952)	1 (0.048)
	Low	1 (0.063)	1 (0.063)	10 (0.625)	4 (0.250)
	Diff	-1(-0.063)	-1(-0.063)	10 (0.327)	-3(-0.202)
26	TOTAL	33*(0.733)	3 (0.067)	2 (0.044)	7 (0.156)
	High	19 (0.905)	1 (0.048)	0 (0.000)	1 (0.048)
	Low	8 (0.500)	1 (0.063)	2 (0.125)	5 (0.313)
	Diff	11 (0.405)	0(-0.015)	-2(-0.125)	-4(-0.265)
27	TOTAL	0 (0.000)	4 (0.089)	37*(0.822)	4 (0.089)
	High	0 (0.000)	1 (0.048)	18 (0.857)	2 (0.095)
	Low	0 (0.000)	2 (0.125)	12 (0.750)	2 (0.125)
	Diff	0 (0.000)	-1(-0.077)	6 (0.107)	0(-0.030)
28	TOTAL	1 (0.022)	30*(0.667)	5 (0.111)	9 (0.200)
	High	0 (0.000)	18 (0.857)	1 (0.048)	2 (0.095)
	Low	1 (0.063)	8 (0.500)	3 (0.188)	4 (0.250)
	Diff	-1(-0.063)	10 (0.357)	-2(-0.140)	-2(-0.155)
29	TOTAL	4 (0.089)	7 (0.156)	30*(0.667)	4 (0.089)
	High	1 (0.048)	0 (0.000)	19 (0.905)	1 (0.048)
	Low	2 (0.125)	6 (0.375)	6 (0.375)	2 (0.125)
	Diff	-1(-0.077)	-6(-0.375)	13 (0.530)	-1(-0.077)
30	TOTAL	27*(0.600)	2 (0.044)	7 (0.156)	9 (0.200)
	High	16 (0.762)	0 (0.000)	1 (0.048)	4 (0.190)
	Low	5 (0.313)	1 (0.063)	6 (0.375)	4 (0.250)
	Diff	11 (0.449)	-1(-0.063)	-5(-0.327)	0(-0.060)