

LERTAP Crib Notes

Based on [Chapter 7](#) of the LERTAP manual by Larry Nelson

Stats1ul Workbook

Res =	A	B	C	D	other	U-L diff.	U-L disc.
Item 1 (Upper)	<u>16</u>	0	0	0		0.59	0.81
Item 1 (Lower)	<u>3</u>	1	3	9			
Item 2 (Upper)	2	0	<u>13</u>	1		0.53	0.56
Item 2 (Lower)	9	0	<u>4</u>	3			
Item 3 (Upper)	0	0	<u>14</u>	2		0.66	0.44
Item 3 (Lower)	0	6	<u>7</u>	3			
Item 4 (Upper)	2	7	0	<u>7</u>		0.34	0.19
Item 4 (Lower)	1	6	5	<u>4</u>			
Item 5 (Upper)	<u>15</u>	0	0	1		0.53	0.81
Item 5 (Lower)	<u>2</u>	1	0	13			

Distractors

In multiple choice exams, the key is to have good items that distract weaker students. In this example, the stronger students are in the upper group, and the weaker students are in the lower group. In Item 1, all of the stronger students (**16**) correctly picked item A. The weaker students picked many different options, and item D seemed like a particularly good distractor.

Discrimination Index (U-L disc)

Measures how well an item fits the idea that it should be neither too easy nor too hard. This number is a percentage and the closer it is to 1, the better.

For Item 1, the **U-L disc** is **.81**, which is good. Of the items in this table from Stats1ul, only Item 4 has poor discrimination, **.19**. In item 4 the strong students were about as distracted as the weak students were.

Good values of U-L disc= Greater than 10-20%

Difficulty Index (U-L diff)

How difficult an item is relates to how many people score correctly on it. In these samples, all the items have reasonable difficulty levels. Good scores for this range from .3-.7 depending on the goals of the test. See (Nelson, 7), for more information.

Good values of U-L diff

Easy Item	..75 and higher
Hard Item	.25 and lower

Stats1f Workbook

Point Biserial [$pb(r)$] and Biserial equivalent [$b(r)$]

Another way to determine item quality is by using correlation. These statistics indicate the relationship between performing well on the test as a whole and choosing the correct answer for this item.

For the most part, in this example, you see positive numbers for the underlined correct items for $pb(r)$ and $b(r)$ and negative numbers for incorrect items, which is a good sign. However, Item 4 is a bad item because both the $pb(r)$ and $b(r)$ which are underlined (Option D) are smaller than the value in an incorrect choice (Option A).

Good Values of $pb(r)$ and $b(r)$

Expert	.30 and above
Good	.20 - .29
Marginal	.09 - .19
Poor Item	below .09

Item	option	wt.	n	p	$pb(r)$	$b(r)$	avg.	z
Item 1	A	1.00	37	0.64	<u>0.54</u>	<u>0.69</u>	31.16	0.45
	B	0.00	1	0.02	-0.38	-1.17	13.00	-2.90
	C	0.00	5	0.09	-0.15	-0.27	26.00	-0.50
	D	0.00	15	0.26	-0.45	-0.61	24.60	-0.76
Item 2	A	0.00	15	0.26	-0.41	-0.56	24.93	-0.70
	B	0.00	0	0.00	0.00	0.00	0.00	0.00
	C	1.00	38	0.66	<u>0.45</u>	<u>0.58</u>	30.74	0.37
	D	0.00	5	0.09	-0.23	-0.42	24.60	-0.76
Item 3	A	0.00	1	0.02	-0.07	-0.20	26.00	-0.50
	B	0.00	8	0.14	-0.36	-0.56	23.88	-0.89
	C	1.00	40	0.69	<u>0.26</u>	<u>0.34</u>	29.95	0.23
	D	0.00	9	0.16	-0.07	-0.11	27.78	-0.17
Item 4	A	0.00	5	0.09	0.21	0.37	32.40	0.68
	B	0.00	19	0.33	-0.04	-0.06	28.37	-0.06
	C	0.00	8	0.14	-0.33	-0.51	24.25	-0.82
	D	1.00	26	0.45	<u>0.06</u>	<u>0.08</u>	29.62	0.17

Stats1ul Workbook

reliability (coefficient alpha):	0.73
index of reliability:	0.86
standard error of measurement:	2.81 (6.2%)

Reliability (Coefficient Alpha)

This statistic measures the degree to which an exam is internally consistent. A higher value on the exam indicates a strong relationship between items on the test. It shows how reliably the same student taking the test again, would achieve the same scores.

Good Values of Coefficient Alpha

High Stakes	.95 and above
Professional	.90
Good Enough	.60-.85
Poor Item	Below .60