

Item Analysis

**Introduction to Item and Test
Analysis**



Aim of Classroom Tests

- Distinguish among students
 - *Assess which students have met the learning goals.*
 - *Identify students who have mastered the learning outcomes.*
 - *Identify students who may be experiencing difficulties.*
- Diagnose learning problems
 - *Identify common errors that students may be making.*
 - *Identify learning goals that students have not achieved.*



Goals of Item Analysis

Item analysis acts as a quality control measure for the test.

- Assess the functioning of the test.
- Determine how well items discriminate among students.
- Diagnose why some items did not work well.
- Assess reliability of the test.

Reliability is necessary but not sufficient for validity.

Creating a database

1	NT Std 1 Maths 2014										
2	ID	@1011	@1012	@1013	@1014	@1015	@1016	@1017	@1018	NUMBER	
3	1	1	1	2	1	2	2	3	2	14	
4	2	1	1	1	0	1	1	3	2	10	
5	3	1	1	1	0	1	1	3	2	10	
6	4	1	1	1	1	0	1	0	0	5	
7	5	1	1	2	1	2	1	3	0	11	
8	6	1	1	2	1	2	1	0	0	8	
9	7	1	1	2	1	2	2	3	2	14	
10	8	1	1	2	1	1	2	3	2	13	
11	9	1	1	2	1	2	1	3	0	11	
12	10	1	1	1	2	2	2	3	2	14	
13	11	0	1	1	1	2	0	0	0	5	
14	12	1	1	0	0	2	0	0	0	4	
15	13	1	1	1	1	2	2	1	2	11	
16	14	0	1	1	0	1	0	0	0	3	
17	15	1	1	2	2	2	2	3	2	15	
18	16	1	1	1	1	2	1	3	2	12	
19	17	1	1	1	1	0	1	0	0	5	
20	18	0	1	0	1	2	1	0	0	5	

Creating a database

NCSE 2016 MATHEMATICS									
stid	gender	subject	item1	item2	item3	item4	item5	item6	item7
340302003	0	G	B	D	B	D	A	C	D
370101063	0	G	B	D	B	A	B	C	B
380203060	1	G	B	D	C	A	B	B	B
110101096	1	G	B	D	B	A	A	B	B
120302021	0	G	C	D	B	C	D	B	C
160010011	0	G	D	D	B	D	B	A	D
330401064	1	G	C	B	B	A	A	B	A
220102024	0	G	B	D	B	D	A	D	D
320201019	1	G	B	D	B	A	C	B	D
370505070	1	G	B	D	B	A	B	B	B
110204007	0	G	B	D	B	A	A	C	D
340302116	0	G	B	B	B	B	C	A	A
340102060	0	G	B	D	B	B	C	A	D
220201189	1	G	C	A	B	A	D	A	C

Information in database

- Name of the test and year group
- Responses given by each student.

Multiple choice- letter selected

Constructed response- score obtained

- Total score of each student on the test.
- Highest and lowest score on the test.
- The students who got the highest and lowest scores.
- Number of students getting each item totally correct, partially correct or incorrect.

Calculations based on information in database

- Percentage of students:
 - *Getting each item correct*
 - *Getting no items correct*
 - *Selecting each option in a multiple choice*
- Average student response for each item
- Dispersion of student scores for each item
- Item difficulty
- Item discrimination
- Reliability of test

Frequency

- Mean

=AVERAGE(D3:D17)

- Standard deviation

=STDEV(C3:C17)

- Count of different scores in a column

=COUNTIF(G3:G17,1)

- Count of total responses on an item

=COUNT(E3:E17)

Frequency output from Lertap

[c18) @10115

<u>Option</u>	n	/900
0	141	15.7%
1	759	84.3%

[c19) @10116

<u>Option</u>	n	/900
0	378	42.0%
1	261	29.0%
2	261	29.0%

[c20) @10117

<u>Option</u>	n	/900
0	126	14.0%
1	644	71.6%
2	114	12.7%
3	16	1.8%

item31 (c34)

<u>Option</u>	n	/10343
A	1,097	10.6%
B	3,117	30.1%
C	5,305	51.3%
D	824	8.0%

item32 (c35)

<u>Option</u>	n	/10343
A	1,303	12.6%
B	3,623	35.0%
C	2,356	22.8%
D	3,061	29.6%

item33 (c36)

<u>Option</u>	n	/10343
A	4,246	41.1%
B	2,040	19.7%
C	1,952	18.9%
D	2,105	20.4%

Frequency output from Lertap

Res =	0	1	2	3	4	other	diff.	disc.	?
ITEM1	24%	<u>76%</u>					0.76	0.33	
ITEM2	23%	<u>77%</u>					0.77	0.40	
ITEM3	47%	<u>53%</u>					0.53	0.57	
ITEM4	32%	<u>68%</u>					0.68	0.61	
ITEM5	8%	<u>10%</u>	<u>82%</u>				0.87	0.41	B

Res =	A	B	C	D	other	diff.	disc.	?
item1	10%	<u>74%</u>	13%	3%		0.74	0.25	
item2	2%	12%	9%	<u>78%</u>		0.78	0.36	
item3	4%	<u>92%</u>	2%	2%		0.92	0.17	
item4	<u>64%</u>	14%	9%	13%		0.64	0.40	
item5	38%	14%	<u>20%</u>	28%		0.20	0.28	D

Frequency output from Lertap

item3 (c6)

option	wt.	n	p	pb(r)	b(r)	avg.
A	0.00	456	0.04	-0.15	-0.33	14.78
<u>B</u>	<u>1.00</u>	<u>9,495</u>	<u>0.92</u>	<u>0.17</u>	<u>0.31</u>	<u>20.14</u>
C	0.00	179	0.02	-0.08	-0.25	15.22
D	0.00	213	0.02	-0.11	-0.31	14.32

item4 (c7)

option	wt.	n	p	pb(r)	b(r)	avg.
<u>A</u>	<u>1.00</u>	<u>6,577</u>	<u>0.64</u>	<u>0.40</u>	<u>0.51</u>	<u>22.13</u>
B	0.00	1,484	0.14	-0.18	-0.28	16.52
C	0.00	943	0.09	-0.17	-0.30	15.87
D	0.00	1,339	0.13	-0.31	-0.50	13.95

item5 (c8)

option	wt.	n	p	pb(r)	b(r)	avg.
A	0.00	3,953	0.38	-0.22	-0.28	17.70
B	0.00	1,416	0.14	-0.11	-0.18	17.66
<u>C</u>	<u>1.00</u>	<u>2,054</u>	<u>0.20</u>	<u>0.28</u>	<u>0.39</u>	<u>24.37</u>
D	0.00	2,920	0.28	0.04	0.05	20.10

Item Difficulty

- The percentage of students who answered an item correctly.

persons getting item correct = p value

of persons attempting item

$R_U - R_L$ = p value

N

- Ranges from 0 to 1
- The higher the difficulty value the easier the item.
- Difficulty index is based both on the item and the sample taking the test.

Item Difficulty output from Lertap

Item difficulty values above 0.35 are considered good

item difficulty bands

.00:

.10:

.20: ITEM13 ITEM17

.30:

.40: ITEM12 ITEM14 ITEM15 ITEM19

.50: ITEM3 ITEM10 ITEM18 ITEM21

.60: ITEM4 ITEM6 ITEM7 ITEM8 ITEM16 ITEM20 ITEM23 ITEM25

.70: ITEM1 ITEM2 ITEM9 ITEM22

.80: ITEM5 ITEM11 ITEM24

.90:

Item Difficulty output from Lertap

item difficulty bands

.00:

.10: item38

.20: item5 item20 item27 item29 item37

.30: item10 item12 item13 item26 item32 item35 item36

.40: item7 item19 item22 item23 item28 item33 item34

.50: item6 item8 item9 item11 item15 item21 item24 item31 item39 item40

.60: item4 item14 item17 item25

.70: item1 item2 item30

.80:

.90: item3 item16 item18

Item Discrimination

- Comparing the number of students with high scores in the test who answered the item correctly with the number of students with low scores on the test who answered the item correctly.

$$\frac{R_U - R_L}{\# \text{ in each group}} = D$$

in each group

- A negative item discrimination value indicates a flawed item.
- A low item discrimination value (0 to 0.2) indicates a need to examine the wording of the item.

Item Discrimination output from Lertap

Item discrimination values above 0.2 are considered appropriate.

item discrimination bands

.00:

.10:

.20:

.30: ITEM1 ITEM21

.40: ITEM2 ITEM5 ITEM6 ITEM11 ITEM14 ITEM18 ITEM22

.50: ITEM3 ITEM8 ITEM12 ITEM13 ITEM16 ITEM17 ITEM19 ITEM20 ITEM24

.60: ITEM4 ITEM7 ITEM15 ITEM25

.70: ITEM9 ITEM10 ITEM23

.80:

.90:

Item Discrimination output from Lertap

item discrimination bands

.00: item12 item38

.10: item3 item11 item16 item18 item34 item37

.20: item1 item5 item6 item7 item28 item30 item36

.30: item2 item13 item17 item22 item24 item25 item26 item27 item29 item31 item35

.40: item4 item8 item10 item14 item15 item21 item32 item33 item39 item40

.50: item9 item20 item23

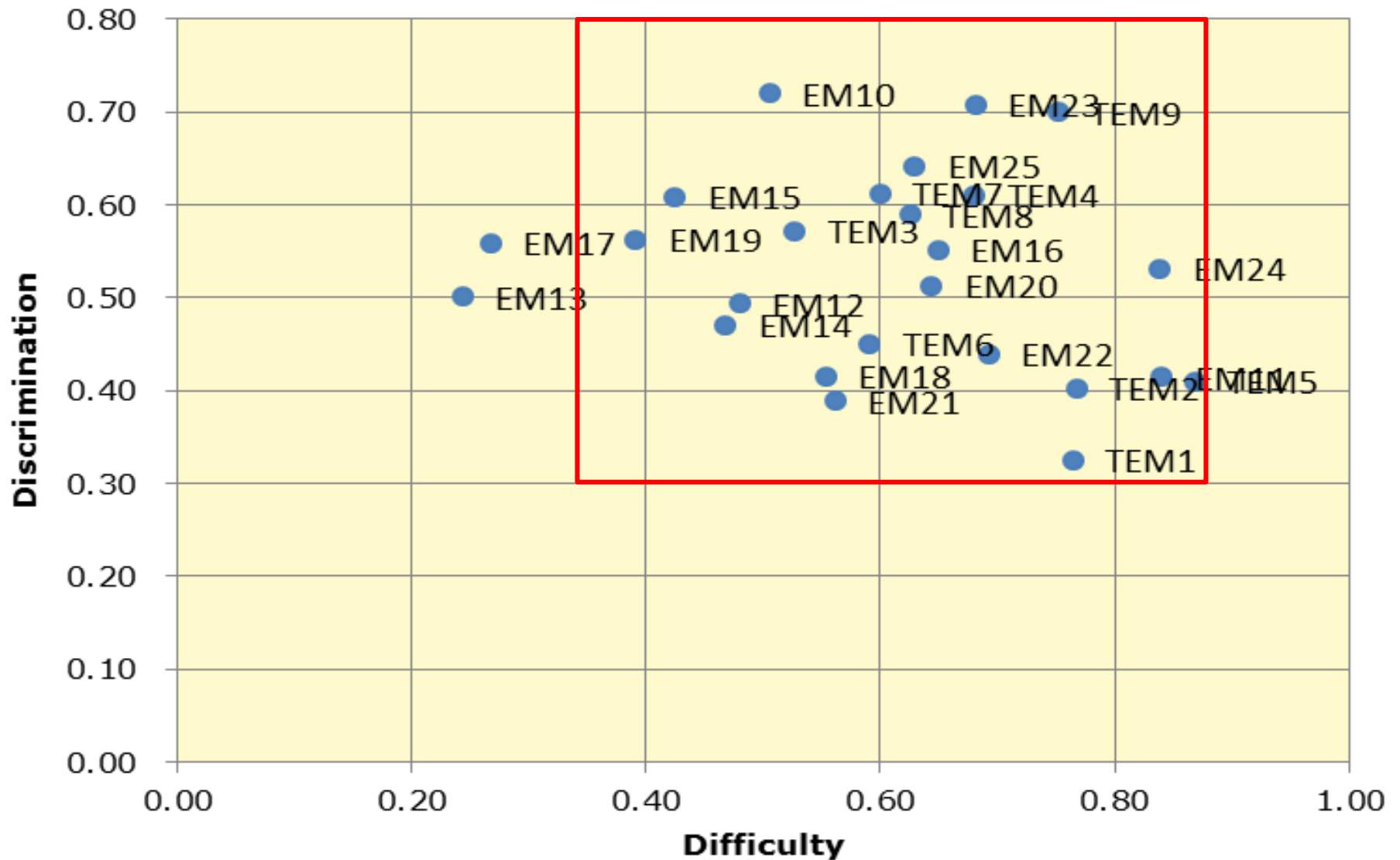
.60: item19

.70:

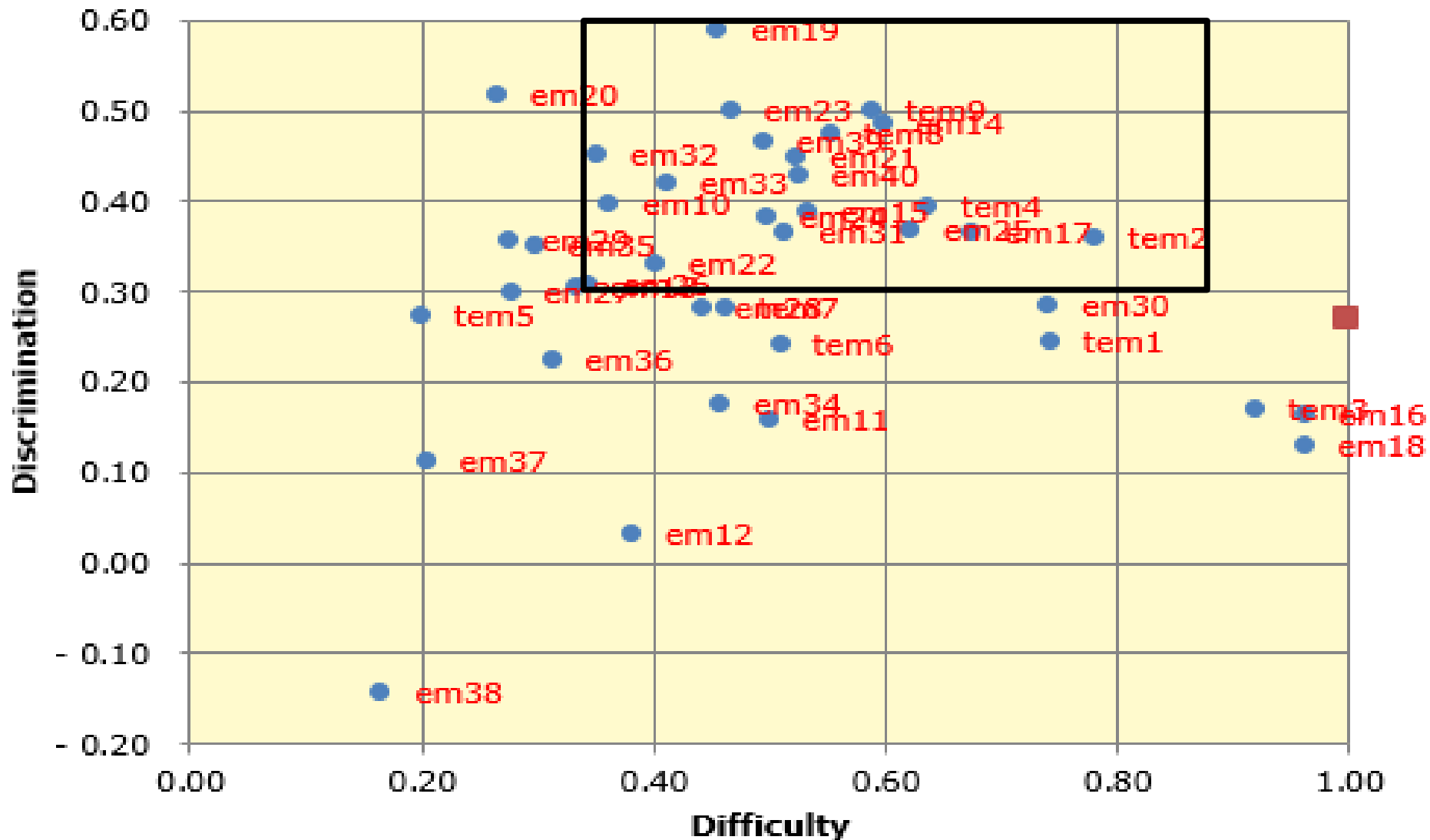
.80:

.90:

Test Analysis output from Lertap



Test Analysis output from Lertap



Test Analysis output by Hand

Discrimination	Difficulty	Questions		
	Hard (0 - 50)	Medium (50 - 85)	Easy (85 - 100)	
Poor (< 0.1)	50	9, 15, 24, 28, 29, 36, 37, 40, 45, 48, 51, 58, 59	27, 30, 32, 49, 55, 57, 60	
Fair (0.1 - 0.3)		21, 38, 41, 47, 54		
Good (> 0.3)	18, 22, 33, 35, 46	3, 4, 12, 17, 31, 34, 39, 44, 52	1, 2, 5, 6, 7, 8, 10, 11, 13, 14, 16, 19, 20, 23, 25, 26, 42, 43, 53, 56	

Test Reliability

- Reliability of a test is the extent to which the test is likely to produce consistent scores.

$$\text{Cronbach } \alpha = \frac{k}{k-1} \left(1 - \frac{\sum_{j=1}^k \text{var}(x_j)}{\text{var}(x_0)} \right)$$

- Ranges from 0 to 1.
- Low reliability means the items were unrelated to each other based on who answered them correctly.
- Classroom tests should have a reliability of 0.7 or more.

Reliability output from Lertap

number of subtest items:	25	
minimum possible score:	0.00	
maximum possible score:	57.00	
reliability (coefficient alpha):	<u>0.91</u>	
index of reliability:	0.95	
standard error of measurement:	3.71	(6.5%)

number of subtest items:	40	
minimum possible score:	0.00	
maximum possible score:	40.00	
reliability (coefficient alpha):	<u>0.85</u>	
index of reliability:	0.92	
standard error of measurement:	2.72	(6.8%)

Reliability output from Lertap

alpha figures (alpha = .9076)

<u>without</u>	<u>alpha</u>	<u>change</u>
ITEM1	0.907	0.000
ITEM2	0.907	-0.001
ITEM3	0.904	-0.003
ITEM4	0.904	-0.003
ITEM5	0.906	-0.002
ITEM6	0.906	-0.002
ITEM7	0.902	-0.006
ITEM8	0.903	-0.004
ITEM9	0.900	-0.008
ITEM10	0.902	-0.006

alpha figures (alpha = .8538)

<u>without</u>	<u>alpha</u>	<u>change</u>
item1	0.852	-0.001
item2	0.850	-0.004
item3	0.853	0.000
item4	0.849	-0.005
item5	0.852	-0.002
item6	0.853	-0.001
item7	0.852	-0.002
item8	0.847	-0.007
item9	0.846	-0.007
item10	0.849	-0.005
item11	0.855	0.001
item12	0.858	0.004