



## APPLICATION OF ITEM ANALYSIS TO EVALUATE THE CHARACTERISTICS OF ITEMS OF THE TEST: QUALITATIVE AND QUANTITATIVE TECHNIQUES

Dr Barkha Devi<sup>1\*</sup>, Ms Shrijana Pradhan<sup>2</sup>, Ms Sandhya Gurung<sup>3</sup>, Ms Doma Giri<sup>4</sup>, Ms  
Nazung Lepcha<sup>5</sup>, Ms Sakila Basnett<sup>6</sup>, Ms Narmaya Chettri<sup>7</sup>

### Abstract

Item analysis is a method for assessing an exam's effectiveness by looking at each of its individual parts or items. It is a procedure of assessing the worth of the test's individual elements and the test as a whole based on student responses to particular test items. It can be utilised to identify elements that are murky or deceptive within a single administration. Additionally, item analysis can be used to pinpoint specific course material passages that need to be emphasised or clarified. A critical stage in the process of standardising summative tests is item analysis. The complexity and internal coherence of the true or false and multiple-choice questions that the instructor produced for the midterm tests given to graduating students of the third year B Sc Nursing at Sikkim Manipal College of Nursing are examined in this study.

Associate Professor<sup>1\*</sup>, Assistant Professor<sup>2,3,4</sup>, Assistant Lecturer<sup>5,6,7</sup>  
<sup>1,2,3,4,5,6,7</sup>Sikkim Manipal College of Nursing, Sikkim Manipal University, Sikkim, India -737102

\* **Corresponding:-** Dr. Barkha Devi,

\*Associate Professor, Sikkim Manipal College of Nursing, Sikkim Manipal University, Gangtok, Sikkim, India, Email-barkha.d@smims.smu.edu.in, Mobile No. 7479278673

**DOI:** - 10.31838/ecb/2023.12.si5.049

## Introduction

Test construction is the set of activities involved in developing and evaluating a test of a certain psychological function. Item analysis is a process that examines student responses to certain test items (questions) in order to assess the effectiveness of the test as a whole and of individual test items (questions). Item analysis can be used to improve items that will be used in upcoming tests as well as remove ambiguous or misleading items from a single test administration. Item analysis is helpful for identifying specific course subject areas that need more attention or clarification as well as for helping teachers create better tests.<sup>1,2,3,4</sup>

The two basic goals of test and item analysis are to enhance instruction and enhance student categorization. The teacher or lecturer can identify the areas of his pupils' weakness and strength and take the necessary action for additional instruction after the test results have been analyzed. Additionally, item analysis can help to enhance the design of exams that will be used to classify students.<sup>2,3,4</sup>

The term "item analysis" refers to a collection of qualitative and quantitative techniques and procedures used to assess the test's items' qualities both before and after their conception and construction. A test's fundamental building block is an item, and an item's analysis tells us how well it performed. Item analysis enables selection and exclusion of things from the test, but it also serves as a tool for item writer improvement. According to Anthony Nitko, some classroom applications of item analysis include determining whether an item works as intended, giving students feedback on their performance, giving teachers feedback on student challenges and suggestions for curriculum improvement, revising assessment tasks, and enhancing item-writing abilities.<sup>5,6,7</sup>

## Item Statistics

On the premise that the overall quality of a test is derived from the quality of its items, item statistics are used to evaluate the performance of individual test items. Item number, item difficulty, and discrimination index are all provided in the item statistics report. Segment statistics, on the other hand, assess the test's overall performance. Data from segment statistics include the mean, standard deviation, mean item difficulty, reliability coefficient, number of items, and standard error of measurement.<sup>7,8</sup>

- **Item number** The question number from the test that was given to the pupils is the item

number. The mean represents the typical or typical student response to a test question. It is calculated by adding together all of the students' points and dividing the result by the total number of pupils.

- **Item difficulty**, the p-value, commonly known as the item difficulty, is the proportion of students who correctly identified the solution. It is frequently known as the item mean. The item difficulty scales from 0 to 100, with a greater value indicating an easier question.
- **Item discrimination** refers to the ability of an item to distinguish students based on how well they comprehend the subject matter being tested is known as item discrimination. Additionally, it shows how effectively the item separates students with high ability from those with low ability. This number lies between 0.0 and 1.0. The item is more discriminating the higher the value.
- **Internal consistency**: The reliability estimate of a test based on a single administration is related to internal consistency. According to Aggarwal in 1986, items with negative discrimination are disqualified from internal consistency tests, which only include items with largely positive correlations to the test score. Due to the different item forms and overall score distributions, the values of the discrimination index will rarely be higher than .50. Item discrimination is often considered to be "good" if the index is higher than .30. Item discrimination is acceptable if the index is between .10 and .30 and poor if it is below.

## Segment Statistics

The dependability, mean score, standard deviation, standard error of measurement, average item difficulty, and number of items are only a few of the data related to segments.<sup>9,10,11</sup>

- **The reliability coefficient** reflects three aspects of the test, including its duration, content, and how the items relate to one another. The reliability of the correlation between the items increases with the ratio of positive relationships and the strength of those relationships. The test's reliability coefficient and item discrimination indices are connected in this way.
- **Mean score** The student's average score within the distribution of scores is indicated by the term "mean score." The measure of central tendency is the mean.
- **Standard deviation (SD)**: The standard deviation (SD) is a measurement of how

widely distributed a student's scores on a given item are. When comparing items with multiple right alternatives or when scale scoring is used, the SD is most useful to utilise. For this reason, the SD is not frequently used to assess exams given in class.

- **Standard error of measurement:** The reliability of the test is closely correlated with the standard error of measurement. It measures the degree of variation in a student's performance brought on by random measurement error. The student's score would alter for a variety of reasons if the test might be given on several days. The scores would have a "normal" shaped distribution for each student. The student's actual score, which measures his level of subject knowledge, is represented by the distribution's mean. The standard deviation, also known as the standard error of measurement, measures how much the student's score could vary from one test administration to the next.
- **Mean item difficulty:** The mean item difficulty for each section is known as the mean item difficulty. Between .40 and .60 is the ideal mean item difficulty level. This score shows that, on average, between 40% and 60% of the candidates successfully responded to the question.
- **Number of items:** The phrase "number of items" refers to the overall quantity of exam questions. The typical number of MCQs in the midterms for MAAP is 30. However, each test should have at least 30 items in order to attain satisfactory exam segment reliabilities.<sup>10,11,12</sup>

## Method

**Objective:** The main objective were

- To deepen knowledge of item analysis and offer practical knowledge for calculating the difficulty (p) and discrimination (d) indices.

## Definitions of item analysis

High (Difficult)	Medium (Moderate)	Low (Easy)
≤ 30%	> 30% AND < 80%	≥ 80%

## Item discrimination index (D)

- The Discrimination Index makes a distinction between students who performed well on the exam and those who performed poorly for each item.
- For each question, deduct the amount of students in the lower group who correctly

- A technique for evaluating how a pedagogical component instructs or how a test question assesses a
- Examining a test's item-by-item responses is known as item analysis. The difficulty of the items should be the main focus of analysis.
- Examining each item on a test to ascertain the percentage of students who chose each response. Can be used to assess a student's strengths and weaknesses; may highlight issues with the validity and potential bias of the test.
- A report of how well each test item performed, typically using traditional statistical criteria like facility and discriminating.

## Purpose

The two basic goals of test and item analysis are to enhance instruction and to better categorise students. The teacher or lecturer can identify the areas of his pupils' weakness and strength and take the necessary action for additional instruction after the test results have been analysed. Additionally, item analysis can help to enhance the design of exams that will be used to classify students.

## Item-difficulty index (p):

This statistic is calculated by calculating the percentage of test takers who correctly respond to the question. the proportion of students who properly answered the question.

The item-difficulty index's formula is

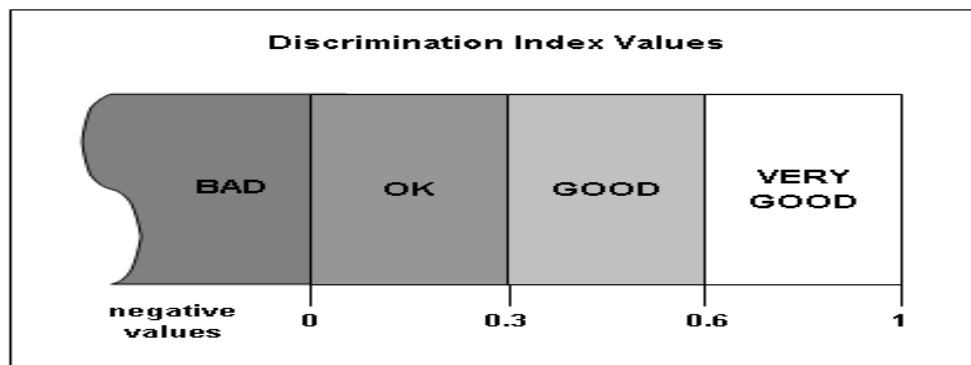
$$p = \frac{N_p}{N}$$

Where:

- $N_p$  indicates the number of test takers in the total group who pass the item, and
- $N$  indicates the total number of test takers in the group.

responded from the amount of students in the higher group.

- Multiply the outcome by the quantity of pupils in each category.
- The Discrimination Index runs from -1 to 1, and it is presented in decimal format



- For exams with a normal distribution, discrimination of 0.3 and above is good; 0.6 and above is very good.
- Values close to 0 mean that most students performed the same on an item.
- The index should never be negative.

The item-discrimination index is calculated using the formula  $d = \frac{U_p - L_p}{U}$

- Where U is the total number of test takers in the upper group,
- $U_p$  and  $L_p$  are the proportions of test takers in the upper and lower o groups who successfully complete the item, respectively.

**Sample of test construction:**

**Student group:** 3<sup>rd</sup> B.Sc. Nursing students  
FULL MARKS- 20.

**TIME- 15 min**

**Part A– 10 Score**

**Direction:** Read the following statements carefully and encircle the alphabet with the correct alternative. Attempt all the questions.

- The following are the importance of ante natal nutrition except:
  - Inadequate nutritional status in pre-conception predisposes to pregnancy complications.
  - Mother’s nutritional status determines the outcome of pregnancy.
  - Mother’s nutritional status determines the nature of delivery.
- Regarding placental transfer of nutrients from the mother to the foetus, it is:
  - Directly from the maternal blood through the placenta into the foetus.
  - From maternal blood to maternal side of the placenta.
  - From maternal blood through the foetal cord into the foetal circulation.

- The most common nutrient transferred through the placenta is:
  - Fats.
  - Carbohydrates.
  - Proteins.
- The pattern of maternal weight gain in pregnancy is as:
  - 1:5:5
  - 2:4:4
  - 1:4:5
- Recommended daily intake of calcium during pregnancy is increase of:
  - 200- 300g.
  - 300- 400g.
  - 400- 500g.
- The increased demand of vitamin D during pregnancy is for:
  - Absorption of calcium & phosphorus.
  - Increased iron absorption.
  - Increased absorption of other vitamins.
- The followings conditions predisposes nutritional risks at the beginning of pregnancy except:
  - Teenage pregnancy.
  - Frequent pregnancies.
  - Elderly pregnancies.
- Nutritional risks during pregnancy includes:
  - Poverty.
  - Low Hb. level.
  - Therapeutic diet.
- General concerns affecting the food intake during pregnancy:
  - Heartburn.
  - Hyperthermia.
  - Dehydration.
- To avoid food related unpleasant effects of iron intake, the precaution which can be taken is:

- Take iron tablets with milk or tea.
- Take iron tablets 1 hr before or 2 hrs after meal.
- Take iron tablets 2 hrs before or 1 hr after a meal.

**Part B – 10 Score**

**Direction:** Read the statements carefully and write true or false against each statement.

- Mother's nutritional status determines the outcome of pregnancy.
- The development of the foetus is directly related to the dietary intake of the mother.
- Cell growth occurs due to the synthesis of proteins in the placenta.

- Strict weight control is necessary during pregnancy to avoid complications.
- An increase of 15 g of calcium is required for a pregnant woman.
- Vit. C is essential for tissue formation and integrity for the foetus.
- Vit.A is needed for cell development and bone growth for the foetus.
- Frequent pregnancy is one of the nutritional risks factors during pregnancy.
- Both inadequate and excessive weight gain are nutritional risk factors at the onset of pregnancy.
- Small frequent meal is an ideal management for heartburn

**Keyanswer sheet: Part A**

Q NO.	ANS.	Q NO.	ANS.	Q NO.	ANS.	Q NO.	ANS.	Q NO.	ANS.
1	c	3	b	5	c	7	c	9	a
2	b	4	a	6	a	8	a	10	b

**Part B**

Q NO.	ANS.	Q NO.	ANS.	Q NO.	ANS.	Q NO.	ANS.	Q NO.	ANS.
1	T	3	F	5	T	7	T	9	F
2	T	4	F	6	T	8	F	10	T

**Blueprint of questions: Part A- questions**

Topics	Knowledge	Understanding	Application	Total marks
• Importance of antenatal nutrition		1 (Q1)		1
• Placental transfer of nutrients.	2 (Q2 & Q3.)			2
• Nutritional demands of pregnancy	3 (Q4, Q5 & Q6)			3
• Risks of poor antenatal nutrition		2 (Q7 & Q8)		2
• Factors contributing to poor nutritional practices & its management	1 (Q9)		1 (Q10)	2
<b>TOTAL</b>	<b>6 (60%)</b>	<b>3 (30%)</b>	<b>1 (10%)</b>	<b>10</b>

**Part B**

Topics	Knowledge	Understanding	Application	Total marks
Importance of antenatal nutrition.		2 (Q1 & Q2)		2
Placental transfer of nutrients.	1 (Q3.)			1
Nutritional demands of pregnancy.	3 (Q5, Q6 & Q7)	1 (Q4)		4
Risks of poor antenatal nutrition.	2 (Q8 & Q9)			2
Factors contributing to poor nutritional practices & its management.			1 (Q10)	1
<b>TOTAL</b>	<b>6 (60%)</b>	<b>3 (30%)</b>	<b>1 (10%)</b>	<b>10</b>

**Lists of students according to marks scored on test:**

RANKS	MARKS SCORED
1.	18.
2.	17.
3.	16.
4.	16.
5.	16.
6.	15.
7.	15.
8.	15.
9.	14.
10.	14.
11.	13.
12.	13.
13.	13.
14.	13
15.	12
16.	12
17.	12
18.	12
19.	11
20.	11
21.	11
22.	10
23.	10

**Calculation of difficulty Index: N= 16**

PART AQ. No.	Difficulty Index%	PART B Q. No.	Difficulty Index%
1.	69	1	88
2	63	2	69
3	50	3	63
4	69	4	81
5	81	5	75
6	75	6	56
7	63	7	50
8	63	8	75
9	69	9	75
10	69	10	50

Each item has a DI above 30%, so all the items are within acceptable for the group.

**Discrimination index:**

PART AQ.No	Discrimination Index %	PART B Q. No	Discrimination Index %
1	0.625	1	0
2	0.5	2	0.62
3	0.75	3	0.5
4	0.625	4	0.375
5	-0.125	5	0.375
6	-0.25	6	0.375
7	0	7	0.5
8	-0.5	8	0.5
9	0.125	9	0.25
10	-0.125	10	0.

All the items are between -1 to 1 , so the questions are good for the group

## Critical evaluation

### Part - A

1. Question No.1, the difficulty index is 69 % whereas discrimination index is 0.625 which means the question is excellent and has high power.
  2. Question No.2, the difficulty index is 63 % whereas discrimination index is 0.5 which means the question is excellent and has high power.
  3. Question No.3, the difficulty index is 50 % whereas discrimination index is 0.75 which means the question is excellent and has high power.
  4. Question No.4, the difficulty index is 69 % whereas discrimination index is 0.625 which means the question is excellent and has high power.
  5. Question No.5, the difficulty index is 81 % whereas discrimination index is - 0.125 which means the question is very easy and poor formation which needs to be revised again.
  6. Question No.6, the difficulty index is 75 % whereas discrimination index is - 0.25 which means the question is poor and which needs to be revised again.
  7. Question No.7, the difficulty index is 63 % whereas discrimination index is 0 which means the question is poorly formed, which needs to be revised again.
  8. Question No.8, the difficulty index is 63 % whereas discrimination index is - 0.5 which means the question is poorly formed, which needs to be revised again.
  9. Question No.2, the difficulty index is 69 % whereas discrimination index is 0.125 which means the question is good and has high power.
  10. Question No.8, the difficulty index is 69 % whereas discrimination index is - 0.125 which means the question is poorly formed, which needs to be revised again.
3. Question No.3, the difficulty index is 63 % whereas discrimination index is 0.5 which means the question is excellent and has high power.
  4. Question No.4, the difficulty index is 81 % whereas discrimination index is 0.375 which means the question is good and has high power.
  5. Question No.5, the difficulty index is 75 % whereas discrimination index is 0.375 which means the question is excellent and has high power.
  6. Question No.6, the difficulty index is 56 % whereas discrimination index is 0.375 which means the question is excellent and has high power.
  7. Question No.7, the difficulty index is 50 % whereas discrimination index is 0.5 which means the question is excellent and has high power.
  8. Question No.8, the difficulty index is 75 % whereas discrimination index is 0.5 which means the question is excellent and has high power.
  9. Question No.9, the difficulty index is 75 % whereas discrimination index is 0.25 which means the question is excellent and has high power.
  10. Question No.10, the difficulty index is 50 % whereas discrimination index is 0 which means the question is poorly formed, which needs to be revised again.
- ✓ In the merit there is huge difference between high group and low group almost in all the question except in question no. 7 in part A and question no. 1 and 10 in part B
  - ✓ Merit of the high group is better than the merit of the lower group in question no 1,2,3,4,9 in part A and question no. 2,3,4,5,6,7,8 and 9
  - ✓ Merit of the lower group is better than the high group for the question no. 5,6,8,9 in part A only.

### Part - B

1. Question No.1, the difficulty index is 88 % whereas discrimination index is 0 which means the question is poorly formed, which needs to be discard.
2. Question No.2, the difficulty index is 69 % whereas discrimination index is 0.62 which

means the question is excellent and has high power.

3. Question No.3, the difficulty index is 63 % whereas discrimination index is 0.5 which means the question is excellent and has high power.
  4. Question No.4, the difficulty index is 81 % whereas discrimination index is 0.375 which means the question is good and has high power.
  5. Question No.5, the difficulty index is 75 % whereas discrimination index is 0.375 which means the question is excellent and has high power.
  6. Question No.6, the difficulty index is 56 % whereas discrimination index is 0.375 which means the question is excellent and has high power.
  7. Question No.7, the difficulty index is 50 % whereas discrimination index is 0.5 which means the question is excellent and has high power.
  8. Question No.8, the difficulty index is 75 % whereas discrimination index is 0.5 which means the question is excellent and has high power.
  9. Question No.9, the difficulty index is 75 % whereas discrimination index is 0.25 which means the question is excellent and has high power.
  10. Question No.10, the difficulty index is 50 % whereas discrimination index is 0 which means the question is poorly formed, which needs to be revised again.
- ✓ In the merit there is huge difference between high group and low group almost in all the question except in question no. 7 in part A and question no. 1 and 10 in part B
  - ✓ Merit of the high group is better than the merit of the lower group in question no 1,2,3,4,9 in part A and question no. 2,3,4,5,6,7,8 and 9
  - ✓ Merit of the lower group is better than the high group for the question no. 5,6,8,9 in part A only.

The theoretical range of reliability coefficients is from zero (perfect reliability) to 1.00 (no reliability). One can be more certain that test results can be used to inform educational decisions if the coefficient is near to 1.0. For almost 95% of the classroom assessments, the range is generally between .50 and .90.

**Reliability of the test construction:**

Odd Scores	Even Scores	R <sub>1</sub>	R <sub>2</sub>	D(R <sub>1</sub> – R <sub>2</sub> )	D <sup>2</sup>
9	9	2	1.5	.5	.25
8	9	5.5	1.5	4	16
8	8	5.5	4	1.5	2.25
8	8	5.5	4	1.5	2.25
9	7	2	8	-6	36
9	6	2	12	-10	100
7	8	9	4	5	25
8	7	5.5	8	-2.5	6.25
7	5	9	14	-5	25
5	7	13	8	5	25
5	6	13	12	1	1
5	6	13	12	1	1
7	4	9	15.5	-6.5	42.25
4	7	15	8	7	49
6	4	11	15.5	-4.5	20.25
3	7	16	8	8	64
				0	415.5

$$r = \frac{1 - 6 \sum D^2}{n(n^2 - 1)}$$

$$= \frac{1 - 6 \times 415.5}{23(529 - 1)}$$

$$= 1 - 0.21$$

= 0.79 (table value at 0.05 at df of 21 is 0.413) so the test is reliable at 0.05 level of significance.

**Conclusion:**

The two basic goals of test and item analysis are to enhance instruction and to better categorise students. The teacher or lecturer can identify the areas of his pupils' weakness and strength and take the necessary action for additional instruction after the test results have been analysed. Additionally, item analysis can help to enhance the design of exams that will be used to classify students. Data from item analysis are speculative. The type and quantity of students being assessed, the teaching methods used, and chance errors all have an impact on these results. Statistics should be kept for each administration of each substance if recurrent use is possible.

**Acknowledgement:** The authors gratefully acknowledge the contribution of all the literatures and of previous study.

**Funding:** None

**Availability of data and materials:** Not applicable

**References**

1. <https://en.Wikipedia.org/wiki/test-construction-strategies>
2. <https://www.sciencedirect.com/topics/social-sciences/test-construction>
3. <https://link.springer.com/referenceworkentry/10.1007%2f978-0-387-79948-3-1257>.
4. American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). Standards for educational and psychological testing. Washington, DC.
5. Aggarwal, Y. P. (1986). Statistical methods, concepts, applications, and computations. New Delhi: Sterling Publication.
6. Boopzthiraj, C. & Chellamani, K. (2013). An analysis of test items on difficulty level and discrimination index in the test for research in education. International Journal of Social Science & Interdisciplinary Research, 2 (2).
7. Lord, F.M. (1952). The relationship of the reliability of multiple-choice test to the distribution of item difficulties. Psychometrika, Vol. 18, pp. 181-194.



8. McAlpine (2002). A summary of methods of item analysis. UK: Computer Assisted Assessment Centre.
9. Mehrens, W.A. & Lehmann, I.J. (1973). Measurement in Education and Psychology. New York: Holt, Rinehart and Winston.
10. Nunally, J.C. (1967). Psychometric Theory. New York: McGraw Hill.
11. Office of Educational Assessment, University of Washington. (2005). Understanding item analysis reports. Retrieved on December 10, 2013 from [http://www.washington.edu/oea/services/scanning\\_scoring/scoring/item\\_analysis.html](http://www.washington.edu/oea/services/scanning_scoring/scoring/item_analysis.html)
12. Franzen, M.D. (2011). Test Construction. In: Kreutzer, J.S., DeLuca, J., Caplan, B. (eds) Encyclopedia of Clinical Neuropsychology. Springer, New York, NY. [https://doi.org/10.1007/978-0-387-79948-3\\_1257](https://doi.org/10.1007/978-0-387-79948-3_1257).

Annexure-I

S.No	Code of student.	1	2	3	4	5	6	7	8	9	10.	1	2	3	4	5	6	7	8	9	10	Total Marks		
1.	A	√	√	√	√	√	√	x	x	√	√	8	√	√	√	√	√	√	√	√	√	10	18	
2.	B	√	√	√	√	x	√	√	x	√	√	8	√	√	√	√	√	√	√	x	√	9	17	
3.	C	√	√	√	√	x	x	x	√	√	x	6	√	√	√	√	√	√	√	√	√	10	16	
4.	D	√	√	√	√	√	√	√	x	√	√	9	√	√	√	√	x	√	x	√	√	10	16	
5.	E	√	x	√	√	√	√	x	√	√	x	7	√	√	√	√	√	√	√	√	x	7	16	
6.	F	√	√	√	√	√	x	√	x	x	√	7	√	√	√	√	x	√	√	√	x	9	15	
7.	G	√	√	√	√	√	x	√	√	x	x	7	√	√	x	√	√	√	x	√	√	8	15	
8.	H	√	√	x	√	√	√	√	x	√	√	8	x	√	√	√	x	√	√	√	x	8	15	
9.	I	x	x	x	x	√	√	√	√	√	x	5	x	x	√	√	√	√	√	√	x	7	12	
10.	J	√	√	x	x	√	√	x	√	x	√	6	√	x	x	√	√	√	x	√	√	x	8	12
11.	K	x	√	x	√	√	√	√	√	√	√	8	√	x	x	x	x	x	x	√	√	3	11	
12.	L	√	√	√	√	x	x	x	x	√	x	5	√	√	x	√	x	x	√	√	√	6	11	
13.	M	√	x	x	x	√	√	√	√	√	√	7	√	x	√	x	√	√	x	x	x	x	4	11
14.	N	x	x	x	x	√	√	√	√	x	√	6	√	√	√	√	x	x	x	x	√	5	11	
15.	O	x	x	x	x	√	√	√	√	√	√	6	√	√	x	x	√	x	x	x	√	x	4	10
16.	P	x	x	x	√	√	√	x	√	x	√	5	√	x	x	√	x	x	√	√	x	√	6	10
	Merit of high group and low group.	8/3	7/3	7/1	8/3	6/7	5/7	5/5	3/7	6/5	5/6	7/7	8/3	7/3	8/5	7/4	6/3	6/2	8/4	7/5	4/4			