

An Analysis of Difficulty Level and Discrimination Indices of Items of Questionnaires on Reproductive Health Awareness of Adolescents

Dr. Deepti Chopra¹, Dr. Shalini Yadava²

¹Assistant Professor, Hindu College of Education, Sonipat, Haryana, India

²Associate Professor, University School of Education, GGSIPU, Delhi, India

Abstract

The study aims to analyze the difficulty level and discrimination index of the items on the questionnaires designed and developed to ascertain the awareness of adolescents about reproductive health. This pilot study was conducted to select the items in the research tools comprising preliminary drafts of two questionnaires having multiple-choice questions on reproductive health. The data was collected through purposive sampling, a total of 109 students of class eight from three private unaided CBSE affiliated schools from Sonipat district of Haryana participated in the study. The data was analyzed for difficulty level and discrimination index of each item on both the tools. The findings of the study led to the selection of the items in the final version of each tool. Those items which did not satisfy the criteria for difficulty level and discrimination index were discarded from the final version of the tool.

Keywords: Difficulty Level, Discrimination Index, Questionnaire, Reproductive Health

Introduction

Research involving elicitation of knowledge, gauging of aptitude, and measuring of perception generally employs a variety of tools including interviews, questionnaires, rating scale, checklist, achievement tests and more. With the intention to ascertain the awareness of very young adolescents regarding reproductive health, the researcher developed questionnaire as a tool as there was a paucity of a standardized instrument for the same. Multiple-choice items were incorporated in the questionnaire, considering the age and other characteristics of the study subjects and the viability of the implementation in the classroom settings. Since this study was a part of a quasi-experimental research work involving an intervention, two parallel forms of questionnaires were constructed, one to be used before and the other to be used after the intervention. The questionnaires comprised of a pool of multiple-choice questions and after determining their validity, the preliminary draft of the same was piloted.

Objectives of the study

The present study was undertaken in order:

1. To find out the difficulty level of the items in the preliminary draft of the tool.
2. To find out the discrimination index of the items in the preliminary draft.

3. To reject the items which do not meet the requirements.
4. To ascertain the validity and reliability of the tool.

Sampling method and Sample

Since the objective of the pilot study was clear, a purposive sampling method was used to collect the data. Permission was sought for three unaided CBSE-affiliated schools in the Sonapat district of Haryana. After data cleaning, the data analysis was performed on a total of 109 students from class eight.

Research Tools

To gain insight about the content on reproductive health for adolescents in the age group of 10-14 years, the report on content analysis of NCERT textbooks with respect to population education (NCERT, 2014) was studied. Apart from this, the Learning Outcomes for Elementary classes published by NCERT were also referred to have a sound knowledge regarding the expected learning outcomes with respect to reproductive health among adolescents (NCERT, 2017). The content of the science books published by NCERT, and other publication houses was analyzed with respect to their content on reproductive health. The analysis revealed that the science curriculum does recognize the curiosity level of elementary school students, but the subject of animal reproduction and adolescence is introduced in class eight. There is no curriculum for adolescence or reproductive health in either class six or class seven. However, a chapter on reproduction in class seven covers the process in plants. The science textbooks of class eight include a chapter 'Reaching the age of Adolescence' which has subtopics - Puberty and changes during puberty, Secondary sexual characteristics, role of hormones, reproductive phase of life in humans, sex determination, hormones other than sex hormones, hormones and life history of insects and frogs and reproductive health and some myths and taboos. Another chapter 'Human reproductive system' in class eight science textbooks explains the functioning of reproductive system in humans. A pool of multiple-choice questions on reproductive health was made based on the content described in the textbooks and selected by the researcher post discussions and brainstorming with science teachers, doctors, counsellors, and eminent educationists. Every multiple-choice question had a stem that presented the main idea or the direction along with a list of four options. The Cambridge dictionary defines awareness as "knowledge that something exists or understanding of a situation or subject at the present time based on information or experience", therefore, the tool comprised of items to assess cognitive skills, specifically knowledge and understanding of the students. Experts in the field were consulted on an initial draft that included 55 items for the pre-test tool and 50 items for the post-test tool. On top of the questionnaires, there was a list of concise instructions that were written in simple terms. The guidelines guaranteed confidentiality and briefly described the goal of the study and the procedure for filling the questionnaires. Additionally, a key for test scoring was also created. Each item in the tool had one correct option. For each correct answer, 1 mark was awarded.

Validity of the Tools

The first drafts of each tool were shown to subject experts to determine face validity. They were asked to review the untitled draft of the tools and provide feedback regarding the intended use of each one. Every input provided a lucid indication of the tools' usefulness, thereby confirming the questionnaires' face validity. The questionnaires were then sent to the experts along with the reference materials and scoring keys for the purpose of content validation. The instruments were updated considering the inputs received,

and the expert opinions, remarks, and recommendations were integrated into the questionnaires which were then revised. Following revision, a few items were removed from the preliminary drafts, and the valid draft of the pre-test tool and the post-test tool comprising 52 items and 48 items respectively were ready for pilot testing.

Data Collection

With proper authorization, the investigator distributed the tool to the students after providing a summary of goals of the study and general guidance on how to complete the questionnaire. After the questionnaires were collected from each class, each student's response sheet was then checked in accordance with the key and finally after weeding out the incomplete answer sheets, data from 109 subjects was coded, and entered in MS Excel for further analysis.

Data Analysis and Interpretation

As per the procedure of item analysis mentioned by Ebel & Frisbie (1991, p.121), the test scores were arranged in rank order. Then the Upper Group (comprising of highest scoring 27 %) and the Lower Group (comprising of lowest scoring 27 %) were identified. The number of respondents from the upper group who selected each response option for each item were counted, and the process was repeated for the lower group.

(a) Item Difficulty Index: According to Matlock Hetzel (1997, p.4), to compute the item difficulty, divide the number of people answering the item correctly by the total number of people answering item.

Item difficulty index is usually denoted as p and is calculated using the formula: $p = R \div T$

where p = Item Difficulty Index

R = Number of Correct responses

and T = Total Number of responses.

Achievement tests are best constructed using items of a moderate difficulty that is within the range of (.30–.70) hence were selected. Items less than 0.30 on the difficulty index were deemed difficult, and more than 0.70 was deemed too easy and were subsequently discarded.

(b) Item Discrimination: The degree to which an achievement test item discriminates between students with high and low achievement is known as its discrimination power. Miller et al. (2012, p. 357) give this formula to compute item discrimination: $D = (RU - RL) \div 0.5T$

where D = Discrimination power

RU = the number of students in the upper group who get the answer right

RL = the number of students in the lower group who get the answer right

T = total number of students included in the items analysis.

Each item in the tools was then evaluated as per the calculated value of index of discrimination as per the criteria shown in table 1.

Table 1: Index of Discrimination

Index of Discrimination	Item Evaluation
0.40 and above	Very good items
0.30 – 0.39	Reasonably good but subject to improvement
0.20 – 0.29	Marginal items, usually needing and being subject to improvement
Below 0.19	Poor items, to be rejected or improved by revision

Note: Reprinted from Essentials of Educational Measurement [e book] (5th ed., p 232), by R.L Ebel and D.A. Frisbie, 1991, Prentice-Hall. Copyright 1991 by Prentice-Hall.

The item wise analysis of 52 items in the pre-test tool in terms of both difficulty index and discriminatory power is provided in table 2. The analysis of Discrimination Power for the pre-test tool indicated that 13 items were very good (item no. 9,10,19,21,23,28,39,41,42,44,47,50,52) and 11 items were reasonably good (item no. 15,17,18,20,25,26,31,32,34,45,51). However, there were 8 items which need improvisation and could be added (item no.3,13,24,37,38,43,48,49) while there were 20 items which were poor and must not be included (item no.1,2,4,5,6,7,8,11,12,14,16,22,27,29,30,33,35,36,40,46).

Table 2: Item wise analysis of Difficulty Index and Discrimination Power of Pre-Test Tool

Item No.	Difficulty Index		Discriminatory Power	
	P	Interpretation	D	Interpretation
1	0.68	Moderate	-0.20	Poor
2	0.91	Easy	0.07	Poor
3	0.52	Moderate	0.23	Needs improvement
4	0.98	Easy	0.03	Poor
5	0.88	Easy	-0.07	Poor
6	0.96	Easy	0	Poor
7	0.93	Easy	-0.10	Poor
8	0.83	Easy	0.13	Poor
9	0.50	Moderate	0.59	Very good
10	0.69	Moderate	0.43	Very good
11	0.93	Easy	0	Poor
12	0.90	Easy	0.07	Poor
13	0.67	Moderate	0.26	Needs improvement
14	0.29	Difficult	0.16	Poor
15	0.30	Difficult	0.33	Reasonably good
16	1	Easy	-0.07	Poor
17	0.69	Moderate	0.39	Reasonably good
18	0.67	Moderate	0.39	Reasonably good
19	0.65	Moderate	0.49	Very good
20	0.68	Moderate	0.36	Reasonably good
21	0.69	Moderate	0.52	Very good
22	0.90	Easy	0.10	Poor
23	0.57	Moderate	0.52	Very good
24	0.61	Moderate	0.26	Needs improvement
25	0.64	Moderate	0.30	Reasonably good
26	0.39	Moderate	0.39	Reasonably good
27	0.98	Easy	0.10	Poor
28	0.66	Moderate	0.56	Very good
29	0.46	Moderate	-0.13	Poor
30	0.93	Easy	0	Poor

Item No.	Difficulty Index		Discriminatory Power	
	P	Interpretation	D	Interpretation
31	0.67	Moderate	0.33	Reasonably good
32	0.68	Moderate	0.33	Reasonably good
33	0.77	Easy	0.07	Poor
34	0.67	Moderate	0.30	Reasonably good
35	0.29	Difficult	0.16	Poor
36	0.92	Easy	0.10	Poor
37	0.69	Moderate	0.20	Needs improvement
38	0.43	Moderate	0.23	Needs improvement
39	0.63	Moderate	0.72	Very good
40	0.95	Easy	0.10	Poor
41	0.70	Moderate	0.46	Very good
42	0.68	Moderate	0.69	Very good
43	0.87	Easy	0.26	Needs improvement
44	0.70	Moderate	0.59	Very good
45	0.44	Moderate	0.33	Reasonably good
46	0.90	Easy	0	Poor
47	0.51	Moderate	0.56	Very good
48	0.52	Moderate	0.23	Needs improvement
49	0.29	Difficult	0.26	Needs improvement
50	0.45	Moderate	0.52	Very good
51	0.64	Moderate	0.39	Reasonably good
52	0.54	Moderate	0.59	Very good

The item-wise analysis of 48 items in the post-test tool in terms of both difficulty index and discriminatory power is provided in table 3. The analysis of Discrimination Power for the post-test tool indicated that 24 items are very good (item no. 3,4,5,6,8,10,11,14,15,18,21,28,29,32,33,35,37,38,39,41,45,46,47,48) and 03 items are reasonably good (item no. 1,43,44). The analysis also revealed that there are 13 items which need improvisation and can be added (item no.2,7,9,13,16,17,19,23,24,25,30,36,40) while there are 08 items which are poor and must not be included (item no.12,20,22,26,27,31,34,42).

Table 3: Item wise analysis of Difficulty Index and Discrimination Power of Post Test Tool

Item No.	Difficulty Index		Discrimination Power	
	P	Interpretation	D	Interpretation
1	0.62	Moderate	0.36	Reasonably good
2	0.66	Moderate	0.29	Needs improvement
3	0.82	Easy	0.54	Very good
4	0.72	Easy	0.50	Very good
5	0.70	Moderate	0.46	Very good
6	0.69	Moderate	0.64	Very good
7	0.68	Moderate	0.29	Needs improvement
8	0.70	Moderate	0.61	Very good

Item No.	Difficulty Index		Discrimination Power	
	P	Interpretation	D	Interpretation
9	0.89	Easy	0.29	Needs improvement
10	0.52	Moderate	0.46	Very good
11	0.62	Moderate	0.43	Very good
12	0.80	Easy	0.14	Poor
13	0.50	Moderate	0.21	Needs improvement
14	0.70	Moderate	0.43	Very good
15	0.69	Moderate	0.68	Very good
16	0.89	Easy	0.21	Needs improvement
17	0.49	Moderate	0.25	Needs improvement
18	0.45	Moderate	0.54	Very good
19	0.67	Moderate	0.29	Needs improvement
20	0.47	Moderate	0.07	Poor
21	0.69	Moderate	0.54	Very good
22	0.40	Moderate	0.18	Poor
23	0.43	Moderate	0.21	Needs improvement
24	0.87	Easy	0.29	Needs improvement
25	0.69	Moderate	0.21	Needs improvement
26	0.52	Moderate	0.14	Poor
27	0.80	Easy	0.14	Poor
28	0.56	Moderate	0.50	Very good
29	0.71	Easy	0.68	Very good
30	0.67	Moderate	0.29	Needs improvement
31	0.89	Easy	0.18	Poor
32	0.46	Moderate	0.61	Very good
33	0.68	Moderate	0.46	Very good
34	0.55	Moderate	0.00	Poor
35	0.57	Moderate	0.57	Very good
36	0.38	Moderate	0.25	Needs improvement
37	0.80	Easy	0.43	Very good
38	0.68	Moderate	0.50	Very good
39	0.63	Moderate	0.57	Very good
40	0.89	Easy	0.29	Needs improvement
41	0.67	Moderate	0.68	Very good
42	0.25	Difficult	0.14	Poor
43	0.30	Difficult	0.32	Reasonably good
44	0.22	Difficult	0.36	Reasonably good
45	0.59	Moderate	0.79	Very good
46	0.81	Easy	0.61	Very good
47	0.66	Moderate	0.54	Very good
48	0.43	Moderate	0.50	Very good

Findings

As shown in table 4, the computation of the Difficulty Index for the pre-test tool indicated that there are 31 items which fall in the category of moderate could be incorporated in the tool. However, there are 4 items which are difficult and 17 items which are easy. On similar lines, the analysis of the post-test tool revealed that there are 33 items which fall in the category of moderate which could be incorporated in the tool. However, there are 3 items which are difficult and 12 items which are easy.

Table 4: Classification of Items in terms of their difficulty index

P	Interpretation	Item no. in Pre-test tool	Item no. in Post-test tool
<.30	Difficult	14,15,35,49	42,43,44
0.30-0.70	Moderate	1,3,9,10,13,17,18,19,20,21,23,24,25,26,28,29,31,32,34,37,38,39,41,42,44,45,47,48,50,51,52	1,2,5,6,7,8,10,11,13,14,15,17,18,19,20,21,22,23,25,26,28,30,32,33,34,35,36,38,39,41,45,47,48
>0.70	Easy	2,4,5,6,7,8,11,12,16,22,27,30,33,36,40,43,46	3,4,9,12,16,24,27,29,31,37,40,46

After analysis of power of discrimination measured as D, 13 items from pre-test tool and 24 items from post-test tool with value of D more than 0.40 were retained as such in the tools. However, 19 items from the pre-test tool and 16 items from post-test tool were modified a little and then incorporated in the tools. Based on low value of Discrimination Index, 20 items from the pre-test tool and 8 items from post-test tool were discarded as their Discrimination Index was less than 0.19 as shown in tabular form in Table 5.

Table 5: Discrimination Index of items in pre-test and post-test tool

D	No. of Items in pre-test questionnaire (%)	No. of Items in post- test questionnaire (%)	Interpretation	Decision
0.40 and above	13(25%)	24(50%)	Very Good	Definitely retain
0.30 – 0.39	11(21.1%)	3(6.2%)	Reasonably Good	Retain
0.20 – 0.29	8(15.3%)	13(27%)	Need Improvement	Improvise and then retain
Below 0.19	20(38.4%)	8(16.66%)	Poor	Discard

Ebel & Frisbie (1991, p. 128) state that the number of items to be included in a test are determined by the amount of time allotted for a test. Since the duration of classes in the school is 35 to 40 minutes, about 30 items were thought to be suitable for incorporation in the final tool. A combined result of both the values of Difficulty Index and Discrimination Power of the pre-test tool consisting of 52 items led to conclusion that 29 items suffice both the criteria so finally item no. 3,9,10,13,17,18,19,20,21,23,24,25,26,28,31,32,34,37,38,39,41,42,44,45,47,48,50,51,52 were incorporated in the final pre-test tool.

On the other hand, a combined result of both the values of Difficulty Index and Discrimination Power of the post-test tool consisting of 48 items led to conclusion that 29 items suffice both the criteria so finally

item no. 1,2,5,6,7,8,10,11,13,14,15,17,18,19,21,23,25,28,30,31,32,35,36,38,39, 41,45,47,48 were incorporated in the final post-test tool. The final tools had 29 items each with the distribution of multiple-choice questions under the different themes as shown below in table 6.

Table 6: Domain wise distribution of items in the Pre-Test and Post-Test Tool

CONTENT AREA	PRE-TEST TOOL		POST-TEST TOOL	
	Knowledge	Understanding	Knowledge	Understanding
Adolescence & Puberty	3	4	3	4
Role of hormones & dev. of sec. sexual characteristics	6	1	6	1
Human Reproductive System	6	2	6	2
Reproductive Health	1	6	1	6
Sub-Total	16	13	16	13
TOTAL ITEMS	29		29	

Reliability of the Tools

Data from 109 students who participated in a pilot study for the pre-test tool were analyzed. The Kuder-Richardson (KR20) coefficient was used to determine the reliability. Both the pre-test and post-test instruments had satisfactory KR-20 coefficient values of 0.78 and 0.79, respectively. This indicates that both instruments are dependable and can be used for research purposes.

Conclusion

The analysis of this pilot study provided a logical basis for selection of items in the final tool. The preliminary draft of both the tools underwent another revision following the results of this pilot testing. This time, the items deemed unfit were removed, and the tool was once more presented to the experts. Not more than 30 items in each tool were incorporated, considering the time allocated for each class. The final tools, the pre-test tool "Questionnaire on Adolescents" and the post-test tool "Post Intervention Questionnaire for Adolescents," each had a total of 29 multiple-choice questions and were ready for implementation. It was made sure that none of the participants from the group trial or pilot study were included in the final research study that followed.

References

1. Cambridge Dictionary (n.d.) Awareness In Cambridge online dictionary retrieved from <https://dictionary.cambridge.org/dictionary/english/awareness>
2. Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of Educational Measurement* (Subsequent). Pearson College Div. <https://ebookppsunp.files.wordpress.com/2016/06/robert-1-ebel-david-a-frisbie-essentials-of-edbookfi-org.pdf>
3. Glen, S. (2017, October 12). *Kuder-Richardson 20 (KR-20) & 21 (KR-21)*. Statistics How To. Retrieved March 4, 2021, from <https://www.statisticshowto.com/kuder-richardson/>
4. Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A Review of Multiple-Choice Item-Writing Guidelines for Classroom Assessment. *Applied Measurement in Education*, 15(3), 309–333. https://doi.org/10.1207/s15324818ame1503_5

5. *Kuder and Richardson Formula 20*. (2021, February 17). Real Statistics Using Excel. <https://www.real-statistics.com/reliability/internal-consistency-reliability/kuder-richardson-formula-20/>
6. Matlock-Hetzel, S. (1997, Jan 23-25). Basic Concepts in Item and Test Analysis [Paper presentation]. *Annual Meeting of the Southwest Educational Research Association*, Austin TX.23 January,1997 <https://files.eric.ed.gov/fulltext/ED406441.pdf>
7. McCowan, R. J., & McCowan, S. C. (1999). *Item Analysis for Criterion-Referenced Tests* (No. ED501716). Center for Development of Human Services; State University of New York (SUNY), Research Foundation. <https://eric.ed.gov/?id=ED501716>