Journal of Education Method and Learnin Strategy
E-ISSN 2986-9129 P-ISSN 3031-9250
Volume 3 Issue 02, May 2025, Pp. 400-411
DOI: https://doi.org/10.59653/jemls.v3i02.1733

Virtual Reality in Biology Assessment: Constructing and Standardizing an Achievement Test for Virtual Rat Dissection

Simmi Sarpal[1*], Anita Nangia[2]
Panjab University, Chandigarh, India[1]
Dev Samaj College of Education, Chandigarh, India[2]
Corresponding Email: simmisarpal@yahoo.com*

Received: 31-03-2025 *Reviewed: 30-04-2025* *Accepted: 17-06-2025*

Abstract

This study presents a detailed process of designing and validating of the biology achievement test for class XI students. The investigator first framed 77 items from the class XI biology NCERT textbook. After writing test items, the researcher consulted subject experts to make sure the items were well-structured, free of language errors, and had the right difficulty level. Incorporating the expert suggestions, a preliminary draft of 80 test items was developed. This preliminary draft was then administered to 100 biology students in private CBSE schools in Chandigarh for item analysis. The difficulty value (DV) and item discrimination (DP) were established using Kelley's (1939) method. Based on item analysis, 38 items were retained in the achievement test. The average item difficulty of the test was calculated as 0.54, and the average item discrimination was 0.49. A high-reliability coefficient of .87 was obtained for the test. The test's validity was supported through expert evaluation and content validation. The results showed that the biology achievement test is both highly reliable and valid for assessing class XI students' understanding of biology.

Keywords: biology achievement test, item analysis, reliability, validity and standardization

Introduction

An achievement test is a knowledge-based instrument educators develop to assess students' success in a subject. It helps evaluate how much knowledge a learner has gained in a specific academic field and offers insights into one's learning capabilities (Bhagat & Baliya, 2016). The achievement test is crucial for educators to assess students' progress, knowledge, skills, or accomplishments in a certain subject area following a training period (Thompson et al., 2021). In educational contexts, these assessments are generally employed to evaluate the extent to which pupils have comprehended and retained the knowledge provided in a particular curriculum or educational program (Hamilton et al., 2021). Anderson (1972) defines an achievement exam as "a collection of questions designed to determine what an individual has learned from instructional exposure." Achievement tests may encompass a diverse array of

subjects, including mathematics, language arts, science, and various other academic fields. Ebel (1965) defines an achievement test as one intended to assess pupils' understanding of knowledge or their competency in certain abilities. (Veronesi, 2017)

Achievement tests can take various forms depending on the subject matter and assessment goals, such as multiple-choice examinations, essay inquiries, practical assessments, and performance evaluations, contingent upon the topic matter being assessed (Reen et al., 2021). Achievement test findings are generally utilized to furnish input to pupils and educators, helping shape future instruction and learning strategies (Ahmad et al., 2020). The researcher designed and standardized an achievement exam to evaluate the biology achievement that resulted from the teaching strategies. The examination was grounded in the CBSE/NCERT syllabus for grade XI.

Requirement for development of Achievement Test

This research aimed to evaluate how effective a virtual dissection approach is when used in biology education for class XI students in comparison to a conventional approach. A 3D virtual model of a rat was developed to augment the learning experience, providing students with engaging and hands on learning experiences, given the anatomical similarities between rats and humans (Romano et al., 2022). The researcher examined current biology accomplishment tests (Sener & Tas, 2017; Singh & Yadav, 2018; Vani & Shabana, 2018; Ahmad et al., 2020; Singh, 2020; Çakır & Görgülü Arı, 2022) but determined that none adequately aligned `with the objectives of the present study. Consequently, the researcher devised an achievement test for biology that aligns with the educational objectives derived from the revised Bloom's taxonomy. The objectives concentrated on recalling, comprehending, and utilizing knowledge from the designated topic specified by CBSE for class XI biology.

Literature Review

Classical Test Theory (CTT) forms the theoretical base for designing and validating achievement-based assessments. Fundamentally, Classical Test Theory assumes that a student's observed test score is a combination of their true score, which reflects their actual ability, and a measurement error. The difficulty index and discrimination index are the main components of the theory. The item difficulty, often represented as the P-value, reflects how many students responded correctly. A well-balanced test includes items with varying levels of difficulty to assess a range of understanding. The discrimination index, or D-value, evaluates how effectively an item differentiates between high-performing and low-performing students by comparing the performance of the top and bottom 27% of students on that item. This helps identify items that are particularly useful for distinguishing between varying levels of ability.

CTT also emphasizes the importance of test reliability and validity. Reliability measures the consistency of the test items, ensuring they collectively assess the intended construct effectively. Validity involves confirming that the test accurately measures the biology content it is supposed to assess. The test construction process under CTT involves designing items that reflect the biology curriculum, creating a test blueprint to ensure comprehensive

coverage of topics, and pilot testing to refine the items based on performance data. Standardization of the test involves establishing norms to interpret individual scores in the context of the broader student population and ensuring comparability across different test versions.

By focusing on item analysis, reliability, and validity, CTT provides a structured and practical approach that provides clear guidelines for creating effective assessment tool.

Objective of the Study

1. To construct an achievement test in biology for class XI students.
2. To standardize the achievement test in biology for class XI students by determining the reliability and validity of the test.

Research Method

The following steps were considered in constructing and validating biology achievement test for class XI students:

1. Planning

Key aspects such as content selection and objective formulation are carefully examined during the planning phase. The test included diverse item types like matching, short answers, multiple-choice, and diagram-based questions. In the specific context of the present study, the achievement test is developed to evaluate students' comprehension of animal biological systems, such as the digestive, respiratory, circulatory, and excretory systems. Content for the test was selected from the Class XI NCERT biology textbook, aligning with the study's objectives. These objectives encompass assessing comprehension, applying knowledge to real-world scenarios, and recognizing the significance of using rats as model organisms, chosen for their physiological resemblance to humans and facilitation of teaching complex biological concepts. Test planning aims to ensure a comprehensive evaluation of students' understanding and appreciation of animal structural organization within the broader context of biology. The study's target population comprises Class XI biology students in CBSE-affiliated private schools in Chandigarh. The biology achievement test includes a variety of item types such as fill in the blanks, match the column, identify a picture, true/false, assertion and reasoning, and multiple-choice questions.

2. Preparation of the first draft of the Achievement Test in Biology

The first draft was developed with a focus on meeting instructional objectives. A total of 77 items were formulated for the preliminary version of the achievement test, taking into account the study's content and objectives. These items comprised various formats, including fill in the blanks, match the column, label a diagram, true/false, assertion and reasoning, and multiple-choice questions. They were structured in alignment with the updated Bloom’s taxonomy framework, covering three dimensions: remembering, understanding, and applying. Subject experts then reviewed the draft to ensure clarity,

relevance, and alignment with learning goals. Their feedback was collected, leading to discussions about potential additions, deletions, or modifications of test items. Following these discussions and incorporating expert suggestions, three items were added (item no. 78, 79, and 80), and adjustments were made to three items concerning language clarity, relevance, and ambiguity (item no. 37, 44, and 73). Consequently, the preliminary draft retained a total of 80 items for the initial trial of the biology achievement test.

3. Pre-try out

The first draft was administered to a sample group comprising 10 Class XII biology students during the pre-tryout phase of developing the achievement test. The main objective of this pre-tryout was to focused on identifying and resolving any linguistic issues within the test items, that students might face in comprehending the test content.

4. Second tryout of achievement test in biology

In the second tryout phase, the initial draft of the achievement test was given to a sample of 100 Class XII biology students from private CBSE schools in Chandigarh, all of whom had already covered the content in their previous class. After completing the test, the students' responses were gathered for subsequent analysis. Student responses were assessed using a standardized predefined scoring key.

5. Item Analysis

Regarding item analysis in educational testing, several methods are available, like Point-Biserial Correlation Coefficient, Item Response Theory, Rasch Model, Distractor Analysis, each with its own strengths and weaknesses. These methods help evaluate the quality of test items to ensure that they effectively measure what they are intended to measure. Among the various methods, Kelley's approach was selected for its ease of use and reliability in educational testing. It provides a reliable measure of item discrimination, helping educators develop tests that accurately and fairly assess student performance.

Based on the tabulated achievement test scores, an item analysis was conducted to optimize the test by selecting the most effective items and eliminating those that performed poorly. The students' total scores were organized in descending order, following Kelley's (1939) dichotomy of the 27% rule. This involved creating an upper group comprising the top 27% of students and a lower group consisting of the lowest 27%. The item analysis focused on two key aspects:

i. Difficulty Index of the Item:

Items that appeared too simple or too challenging for most students were eliminated to maintain balance.

ii. Discriminating Power of the Item:

This analysis was conducted to assess the effectiveness of items in distinguishing between high and low performers. Items that failed to discriminate effectively between these groups were considered for review. This meticulous item analysis ensured the refinement of the

achievement test by retaining items that balanced difficulty and effectively discriminated between students with varying levels of achievement.

D.V. and D.P. were computed according to the following formulae:

$$D.V. = \frac{RU + RL}{N}$$

$$D.P. = \frac{RU - RL}{N/2}$$

Where

RU = Number of right responses in the upper group.

RL = Number of right responses in the lower group.

N = Total number of students in both the groups.

The D.V. and D.P for each item was calculated and is given in the following Table 1.

Table 1. Difficulty value (D.V) and Discriminatory power (D.P) of the Items in Biology Academic Achievement Test

Item No.	RU	RL	D.V.	D.P.	Remarks
1	27	25	0.962	0.074	Rejected
2	14	12	0.481	0.074	Rejected
3	26	20	0.851	0.222	Accepted
4	23	8	0.574	0.555	Accepted
5	24	17	0.759	0.259	Accepted
6	19	14	0.611	0.185	Rejected
7	18	23	0.759	-0.185	Rejected
8	26	21	0.870	0.185	Rejected
9	19	14	0.611	0.185	Rejected
10	11	8	0.351	0.111	Rejected
11	24	21	0.833	0.111	Rejected
12	16	6	0.407	0.370	Accepted
13	12	6	0.333	0.222	Rejected
14	6	5	0.203	0.037	Rejected
15	17	11	0.518	0.222	Accepted
16	27	23	0.925	0.148	Rejected
17	5	6	0.203	-0.037	Rejected
18	22	6	0.518	0.592	Accepted
19	24	16	0.740	0.296	Rejected
20	27	24	0.944	0.111	Rejected
21	26	19	0.833	0.259	Rejected
22	27	11	0.703	0.592	Accepted
23	17	8	0.462	0.33	Accepted

24	6	1	0.129	0.185	Rejected
25	26	24	0.925	0.074	Rejected
26	19	7	0.481	0.44	Accepted
27	26	23	0.925	0.148	Rejected
28	16	2	0.333	0.518	Accepted
29	25	8	0.611	0.629	Accepted
30	24	7	0.574	0.629	Accepted
31	17	6	0.425	0.407	Accepted
32	25	9	0.629	0.592	Accepted
33	25	12	0.685	0.481	Accepted
34	25	10	0.648	0.555	Accepted
35	25	8	0.611	0.629	Accepted
36	23	5	0.518	0.666	Accepted
37	23	9	0.592	0.518	Accepted
38	21	7	0.518	0.518	Accepted
39	22	6	0.519	0.592	Accepted
40	8	6	0.259	0.074	Rejected
41	25	24	0.907	0.037	Rejected
42	6	0	0.111	0.222	Rejected
43	25	12	0.685	0.481	Accepted
44	18	10	0.518	0.29	Accepted
45	8	4	0.222	0.148	Rejected
46	5	10	0.277	-0.185	Rejected
47	19	4	0.425	0.555	Accepted
48	14	21	0.648	-0.259	Rejected
49	13	15	0.518	-0.074	Rejected
50	14	17	0.574	-0.111	Rejected
51	19	18	0.685	0.037	Rejected
52	16	15	0.574	0.037	Rejected
53	22	13	0.648	0.333	Accepted
54	16	9	0.462	0.259	Accepted
55	16	11	0.5	0.185	Rejected
56	20	11	0.574	0.33	Accepted
57	24	24	0.888	0	Rejected
58	4	8	0.222	-0.148	Rejected
59	18	7	0.462	0.407	Accepted
60	19	2	0.388	0.629	Accepted
61	4	3	0.129	0.037	Rejected
62	20	10	0.555	0.370	Accepted
63	24	7	0.574	0.629	Accepted
64	7	4	0.203	0.111	Rejected
65	12	0	0.222	0.444	Accepted
66	11	0	0.203	0.407	Rejected
67	12	0	0.222	0.444	Rejected
68	9	0	0.166	0.333	Rejected
69	9	6	0.277	0.111	Rejected

70	5	0	0.092	0.185	Rejected
71	5	1	0.111	0.148	Rejected
72	17	1	0.333	0.592	Accepted
73	5	0	0.092	0.185	Rejected
74	23	4	0.5	0.703	Accepted
75	3	0	0.055	0.111	Rejected
76	24	12	0.666	0.444	Accepted
77	23	24	0.870	-0.037	Rejected
78	23	12	0.648	0.407	Accepted
79	24	4	0.518	0.740	Accepted
80	21	3	0.444	0.666	Accepted

Result

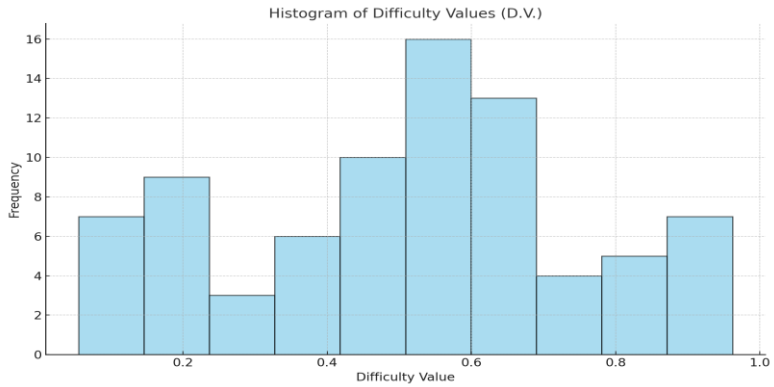

Figure 1: Histogram of Difficulty Values

Figure 1 displays the frequency distribution of difficulty values (D.V.) ranging from 0 to 1. It shows that most values are concentrated around 0.6, with fewer entries towards the extremes.

The histogram illustrates the frequency distribution of discrimination power (D.P.) values, ranging from approximately -0.2 to 0.7. It shows that most values are centered around 0.2, with fewer entries at the extremes (Figure 2).

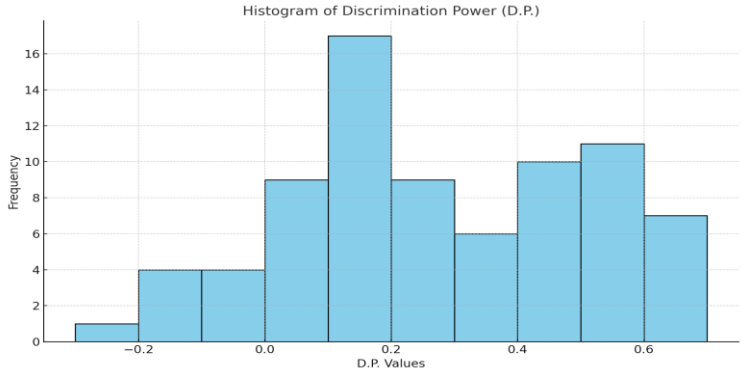

Figure 2: Histogram of Discriminatory Power

The items were chosen based on Ebel's 1966 criteria for Difficulty Value (D.V.) and Discriminatory Power (D.P.), as detailed in the tables presented as Table 2 and Table 3 below.

Table 2 Distribution items of achievement test as per Ebel's Index of Difficulty Value

Index of Difficulty Value	Item evaluation
0.67 and above	Poor Items
Between 0.60 to 0.67	Needs Improvement
Between 0.20 to 0.59	Very Good Items
<0.20	Very Poor

Table 3 Distribution of items of achievement test as per Ebel's Discriminating Power

Discriminatory Power	Remarks
0.40 and above	Very Good Items
Between 0.30 and 0.39	Reasonably Good
Between 0.20 and 0.29	Needs Improvement
< 0.19	Very Poor

Items deleted after item analysis: 1, 2, 6, 7, 8, 9, 10, 11, 13, 14, 16, 17, 19, 20, 21, 24, 25, 27, 40, 41, 42, 45, 46, 48, 49, 50, 51, 52, 55, 57, 58, 61, 64, 66, 67, 68, 69, 70, 71, 73, 75 and 77. Items modified after item analysis: 3, 5, 15, 65, and 80.

After item analysis, the items with average difficulty value and appropriate discrimination power were retained in the test. Out of 80 items 42 items were rejected and 38 items were retained for the final draft of the achievement test given in Table 4.

Table 4 Number of Items after Item Analysis of First Draft of Achievement Test

Total Items	Items Accepted	Items Rejected
80	38	42

The final draft of the achievement test had 38 items, including multiple choice questions: fill in the blanks, complete with hint, and match the column, true false, diagram question and short answer type questions. The blue print for the final draft of the achievement test based on the revised Bloom's taxonomy is given in Table 5.

Table 5 Blueprint of the Final Draft of Achievement Test for Biology

Objectives Content	Remembering	Understanding	Applying	Total
Morphology of Rat	2	-	-	2
Digestive System of Rat	3	9	1	13
Respiratory System of Rat	1	2	5	8
Circulatory System of Rat	3	2	1	6
Excretory System of Rat	3	2	1	6
Reproductive System of Rat	1	-	1	2
Dissection Process	-	1	-	1
Total	13	16	9	38

6. Standardization of Achievement Test

The achievement test was further standardized by reliability and validity.

Reliability

When discussing achievement tests, reliability is defined as the degree to which scores remain constant and consistent over time and administrations. When administered under consistent conditions to the same participants, a valid test ought to produce the same findings every time. Reliability was determined using a test-retest method, assessing score consistency over time. A group of students took the exam first, and then another group took it again fifteen days later. In order to determine the Pearson's coefficient of correlation, the following formula was used:

$$\mathbf{r} = \frac{\mathbf{n}(\sum \mathbf{xy}) - (\sum \mathbf{x})(\sum \mathbf{y})}{\sqrt{[\mathbf{n}\sum \mathbf{x^2} - (\sum \mathbf{x})^2][\mathbf{n}\sum \mathbf{y^2} - (\sum \mathbf{y})^2]}}$$

Where,

r = Pearson correlation coefficient

x = Scores in the first set of data

y = Scores in the second set of data

n = Total number of scores.

A correlation of 0.87 was determined between the scores. Therefore, the test's reliability coefficient was 0.87. The biology achievement test is quite reliable, according to this value.

Validity

How well a test measures its intended constructs is called validity. According to this study, a legitimate achievement test in biology should accurately evaluate students' grasp of biological principles, concepts, and abilities without introducing biases or irrelevant aspects. This study primarily focused on validating whether the test content accurately represents the intended curriculum. In the case of a biology achievement test in particular, this means that the questions are structured to cover the most important ideas covered in the course. The same experts who had previously reviewed the test also reviewed it to determine its content validity. Interestingly, the test items were highly congruent throughout all experts, confirming that they were in line with the desired content. This confirms that the test accurately measures the desired biological knowledge and abilities and is valid.

Percentile norms and stanine norms were established which are given in Tables

PERCENTILE NORMS

Percentile	Raw Scores	Percentile	Raw Scores	Percentile	Raw Scores	Percentile	Raw Scores
P99	79	P75	48	P50	41	P25	32
P95	55	P70	47	P45	40	P20	30
P90	52	P65	46	P40	38	P15	28
P85	50	P60	45	P35	36	P10	27
P80	49	P55	43	P30	34	P5	22

STANINE NORMS

Sr. No.	Stanine Scale	Percentile	Raw Score	Interpretation
1	9	P99	79	Outstanding
2	8	P90	68	Very High
3	7	P80	64	High
4	6	P65	60	Above Average
5	5	P50	55	Average
6	4	P30	50	Lower Average
7	3	P15	45	Below Average
8	2	P5	40	Low
9	1	P1	16	Very Low

Scoring:

For each correct answer one mark was awarded and for wrong answer zero mark was given for question no. 1 to 33, two marks were given for question no. 34 to 37 and 4 marks for question no. 38.

Discussion

The biology achievement test, developed and standardized by the researcher for Class XI, has demonstrated reliability and validity, as evidenced by the reliability coefficient and item analysis results. The item analysis procedure revealed that the difficulty values of the items fall within a moderate range, reflecting a balanced challenge level for students. Additionally, the discriminatory power of the items also falls within a moderate range, suggesting that the assessment successfully differentiates learners based on their academic performance levels. It was decided that those items with a difficulty value (D.V.) falling below 0.40 or exceeding 0.60 were rejected, and only the items with D.V. scores ranging from 0.40 to 0.60 were selected for the test. In line with Ebel's (1966) criterion, items were also selected based on their Discriminatory Power (D.P.). Those items with a D.P. of 0.40 or higher were retained, while items with a D.P. below 0.40 were not included in the final test. Careful selection of test items guarantees a suitable difficulty level that aligns with the students' capabilities.

Conclusion

This assessment tool, featured in the researcher's doctoral study, is highly recommended for use by teachers, educators, and researchers to evaluate the academic achievements of Class XI biology students. Given its high reliability and validity, it is recommended that educators, teachers, and researchers utilize this test as a valuable tool for assessing the achievement of Class XI biology students. This endorsement underscores the test's potential to provide accurate and consistent measurements, positioning it as a useful instrument for academic assessment and further education studies.

References

Ahmad, S., Sultana, N., & Jamil, S. (2020a). Considerations for constructing and validating biology achievement test at secondary level. *US-China Education Review B, 10*(1). https://doi.org/10.17265/2161-6248/2020.01.002

Anderson, L. W., & Krathwohl, D. R. (2001). A taxonomy for learning, teaching, and assessing: A Revision of Bloom's Taxonomy of Educational Objectives. New York: Longman.

Anderson, R. C. (1972). How to Construct Achievement Tests to Assess Comprehension. *Review of Educational Research, 42*(2), 145-170. https://doi.org/10.3102/00346543042002145

Bhagat. P., & Baliya, J.N. (2016). Construction and Validation of Achievement Test in Science. *International Journal of Science and Research, 5*(6), 2277-2280.

Ebel, R. L. (1965). Measuring Educational Achievement. Englewood Cliffs, N.J.: Prentice-Hall.

Hamilton, D., McKechnie, J., Edgerton, E., & Wilson, C. (2021). Immersive virtual reality as

a pedagogical tool in education: a systematic literature review of quantitative learning outcomes and experimental design. *Journal of Computers in Education*, *8*(1). https://doi.org/10.1007/s40692-020-00169-2

Malay, C.A., Fabia, J.N.V. & Santillan, E.P. (2023). Construction and Validation of Biology Assessment Test (BAT) for Junior High School Students. *European Journal of Theoretical and Applied Sciences*, 1(3), 134-139. DOI: 10.59324/ejtas.2023.1(3).13

Reen, F. J., Jump, O., McSharry, B. P., Morgan, J., Murphy, D., O'Leary, N., O'Mahony, B., Scallan, M., & Supple, B. (2021). The Use of Virtual Reality in the Teaching of Challenging Concepts in Virology, Cell Culture and Molecular Biology. *Frontiers in Virtual Reality*, *2*. https://doi.org/10.3389/frvir.2021.670909

Romano, M., Bravin, A., Wright, M. D., Jacques, L., Miettinen, A., Hlushchuk, R., Dinkel, J., Bartzsch, S., Laissue, J. A., Djonov, V., & Coan, P. (2022). X-Ray Phase Contrast 3D Virtual Histology: Evaluation of Lung Alterations After Microbeam Irradiation. *International Journal of Radiation Oncology Biology Physics*, *112*(3). https://doi.org/10.1016/j.ijrobp.2021.10.009

Singh, A., & Yadav, D. (2018). Construction and standardization of achievement test in Biology. *International Journal of Research in Social Science, 8*(3), 18-27.

Singh, D. K. (2020). Construction and Standardization of Biology Achievement test for Higher Secondary Students. *Advances in Chemical and Applied Sciences, 3*.

Sener, N., & Tas, E. (2017). Developing achievement test: A research for assessment of 5th grade biology subject. *Journal of Education and Learning, 6*(2), 254. https://doi.org/10.5539/jel.v6n2p254

Thompson, M., Uz-Bilgin, C., Tutwiler, M. S., Anteneh, M., Meija, J. C., Wang, A., Tan, P., Eberhardt, R., Roy, D., Perry, J., & Klopfer, E. (2021). Immersion positively affects learning in virtual reality games compared to equally interactive 2d games. *Information and Learning Science*, *122*(7–8). https://doi.org/10.1108/ILS-12-2020-0252

Vani, T. & Shabana. (2018). Construction and Standardization of an Achievement Test of Biological Science for class IX Students. *Int. J. Rev. and Res. Social Sci, 6(*11), 12-14

Veronesi, G. (2017). ES 05.04 Minimally Invasive Surgery for Lung Cancer, including Robotics. *Journal of Thoracic Oncology*, *12*(11). https://doi.org/10.1016/j.jtho.2017.09.123