



ASSESSING SCIENCE TEACHERS' KNOWLEDGE ABOUT INSTRUCTIONAL STRATEGIES: DEVELOPMENT AND VALIDATION OF A MCQS TEST

Muhammad Afzal¹, Dr Khalid Saleem², Dr Syed Abdul Waheed³

Muhammad Afzal

Ph.D. Scholar, University of Okara

Email: afzalfg@ymail.com

Dr. Khalid Saleem

Corresponding Author

Associate Professor, Department of Teacher Education, University of Okara, Pakistan

Email: khalid.saleem@uo.edu.pk

Dr. Syed Abdul Waheed

Assistant Professor, Department of Educational Studies, University of Okara, Pakistan

E-mail: s.a.waheed@uo.edu.pk

Abstract

Effective science teaching requires strong critical-thinking, problem-solving, communication, collaboration, and creative skills. Effective science teaching can only be ensured if the teacher, the key pivot of the change, is well-acquainted with content as well as methodology. Therefore, classroom practices are becoming more focused on the delivery of instruction through innovative instructional strategies. On the other hand, instructional planning and strategies are the key facets of quality teaching. Most of the scales developed for measuring science teachers' knowledge about instructional strategies had rarely focused on these aspects. This study aims to design and validate an MCQ test to evaluate the instructional strategy knowledge of science teachers. The test, comprising 44 items, was administered to 200 science teachers at the secondary school level. To establish the validity and reliability of the MCQ test, the discrimination index, difficulty index, distractor efficiency, and the Kuder-Richardson Formula 20/20 (ρ_{KR20}) were employed. The final MCQs test consisted of 31 items. The ρ_{KR20} score obtained was 0.85, showing that the test is highly reliable.

Keywords: *Discrimination index (Dis I), The distractor efficiency (DE), Instructional strategies, Multiple-choice items, Difficulty index (DI).*

Introduction

Teaching is a complicated process because it is both an art as well a science (Marzano, 2007, 2017) which in one way or another other is dependent on teachers' instructional strategies. Instructional knowledge i.e., knowledge of instructional strategies of teachers is an indicator of their efficiencies. It refers to the specified knowledge that creates effective teaching and a conducive learning environment (Guerriero, 2014). Various studies have highlighted the fact that an effective teacher has a pivotal role in student learning (Darling-Hammond, 2000; Marzano, 2007). Additionally, Marzano (2007) explains that effective teachers select strategies based on their students' needs and interests. Teachers use appropriate instructional strategies to make their work easier, according to their experiences, personal preferences and norms (Watson, 2003). Moreover, students' level of academic achievement, in one way or the other, depends upon the effectiveness of teachers' instructional strategies and vice versa (Kimani, Kara, & Njagi, 2013).



Teachers have to possess an awareness of effective ways to transfer knowledge rather than just pass knowledge to the learners. Such teachers are capable to encourage, motivating, understanding, and caring to inspire students and contribute to students' achievement (Turkmen, 2009). A teacher with good instructional knowledge, dispositions, and skills can provide a learning environment that supports students' intellectual, social, and personal development (NCARE, 2002). Therefore, the teacher has to know the specific function of instructional strategies and the context in which these strategies are being applied (Lohse-Bossenz, Kunina Habenicht, & Kunter, 2013).

Unlike other subjects, science teaching can help students in retrieving new ideas. It can lead them to observation-based discoveries, critical thinking and acting rationally. Teachers should possess theoretical as well as practical knowledge and abilities to teach science. In science teaching, the focus is to develop problem-solving abilities, critical thinking, and skills among students. All the teaching methods are not equally beneficial for all the students. Some strategies may attract a few students and distract several others and vice versa (Gözütok, 2000). Therefore, for science teachers, it is important to use teaching methods/strategies that can facilitate the majority of students despite their varied needs, interests and abilities.

In collaboration with UNESCO, the Government of Pakistan (2009) designed National Professional Standards for Teachers (NPSTs) to enhance the teaching-learning process and improve the overall quality of education. These standards help teachers to understand the instructional planning comprising of subject matter knowledge, students' needs, community preferences, objectives of the curriculum, and adopting effective teaching strategies to promote students' problem-solving skills and critical thinking. The role of teachers' evaluation is important in effective teaching. It helps in assessing teachers' strengths and weaknesses, enabling continued development (Santiago and Benavides 2009). Therefore, it becomes important to use appropriate tests to gauge teachers' professional knowledge (Großschedl, Welter, & Harms, 2019). Therefore, the study focused on designing and validating an assessment tool to quantify science teachers' instructional strategy knowledge.

Among various types of objective tests, MCQ tests are regarded as well-structured instruments for evaluating knowledge, and their overall quality relies on the quality of each individual item (Beullens et al. 2002). These tests, aligned with the Bloom's cognitive domain (Mehta & Mokhasi, 2014), are designed to evaluate respondents' knowledge by revealing their strengths and weaknesses in a specific subject area (Gierl et al. 2017) and to measure more complex cognitive processes as well (Kıyak et al., 2022). Knowledge assessments are being taken using the MCQ tests as they are capable of measuring vast content within a limited period of time (Christian et al. 2017). However, to achieve the desired outcomes, the test should be developed by skilled persons otherwise, it can be confusing for the test-takers. It can be either too easy or too difficult and fail to accomplish the desired outcomes. If the alternatives in the MCQs test are not according to standardized criteria, it can lessen recalling or comprehension. Ultimately, it leads to guessing (Palmer & Devitt, 2007; Kuechler, & Simkin, 2010; Haladyna, 2013).

Among different types of objective type tests, the MCQs tests are considered well-organized instruments, and their quality depends upon the quality of each MCQs item (Hingorjo & Jaleel, 2012). After conducting the test, a particular procedure is followed to analyze every item individually (DiBattista & Kurzawa, 2011). The test items can be validated using various parameters which include the the discrimination index (Dis I), difficulty index (DI), distractor efficiency (DE) and the Kuder-Richardson Formula 20 (Earnest et al. 2018). These parameters are described in detail in the methodology section.

Researches on the Assessment of Teachers' Instructional Knowledge

Teachers' instructional knowledge plays a vital role to understand the teaching-learning process. It also helps teachers in decision-making about lesson design or on-the-spot judgments in the classroom (Guerriero, 2014). Studies have corroborated that teachers' knowledge about instructional strategies improves the teaching practices in the classroom and ultimately it promotes students' learning. It suggests that teachers' knowledge of instructional strategies is essential to improve teachers' competencies to create an effective learning environment (Meyen & Greer, 2009; Ruys, Van Keer, & Aelterman, 2012).

Several studies have been conducted to assess teachers' pedagogical knowledge with different measuring instruments. Generally, it is perceived that pedagogical content knowledge (PCK) includes at least two major facets—knowledge of effective instructional strategies and knowledge of how students comprehend content (Lee & Luft, 2008; Depaepe, Verschaffel, & Kelchtermans, 2013). Voss et al. (2011) used an MCQs test to assess knowledge of teaching methods. The Teacher Knowledge Survey was developed in the Centre for Education Research in Innovation (CERI) to measure the overall pedagogical knowledge of teachers, which is needed for effective teaching in a conducive learning environment (Guerriero, 2017).

The instrument "Innovative Teaching for Effective Learning Teacher Knowledge Survey (ITEL TKS)" was developed by Sonmark, K. *et al.* (2017) to assess teachers' pedagogical knowledge in the domains of instruction and assessment. Moreover, the assessment tool "Measuring the professional knowledge of pre-service mathematics and science teachers" was developed for measuring the professional knowledge of pre-service teachers of mathematics, physics, chemistry and biology (Kleickmann et al., 2014; Großschedl, Harms, Kleickmann, & Glowinski, 2015)

All the instruments mentioned above were designed to assess various dimensions of teachers' pedagogical knowledge. However, they do not encompass the different aspects of instructional strategies, i.e. orientation about instructional strategies, knowledge of lecture/discussion strategies, knowledge of problem-solving strategies, knowledge of demonstration strategies, knowledge of questioning/inquiry strategies, and knowledge of cooperative learning strategies as defined by the NPSTs. Thus, this study emphasizes the development of a scale for assessing science teachers' knowledge of instructional strategies.

Materials and Methods

The study at hand aimed to develop a scale for assessing knowledge of science teachers' instructional strategies and the positivist paradigm was considered appropriate to be followed to conduct the study. As a quantitative approach, it focuses on predicting, controlling, and generalizing findings via surveys, questionnaires, or experimental methods. "Positivism is concerned with uncovering truth and presenting it by empirical means" (Henning, Van Rensburg & Smit, 2004, p.17). This study was a descriptive one and the survey technique was applied to collect data.

After a detailed review of the related literature and NPSTs (Government of Pakistan, 2009) a MCQ test was developed to cover the six commonly used instructional strategies being practised by the science teachers viz. orientation about instructional strategies, lecture/discussion strategies, problem-solving strategies, demonstration strategies, questioning/inquiry strategies and cooperative learning strategies. Each test item included four alternatives, consisting of one correct answer and three distractors. Before administering the test, both face validity and content validity were established by educational experts.

All male and female science teachers teaching in urban and rural public secondary schools within a district of Pakistan during the 2019–20 session formed the target population



of the study. The MCQs tests were distributed to randomly selected 232 science teachers, and 200 dully filled scales were returned. The Dis I, the DI, the DE as well as Kuder-Richardson Formula 20 were used to calculate the reliability and internal validity of the developed test.

The Difficulty Index

The quality of a multiple-choice item can be effectively evaluated using the Difficulty Index (DI) and Discrimination Index (Dis I) (Sutopo et al, 2024). The difficulty index (*DI*) measures the proportion or percentage of the respondents who correctly answer the given test item. Its range is from 0% to 100%. The higher value of the DI indicates that most of the examinees respond to the item correctly, and it was thus rated an easier item (Malau-Aduli & Zimitat 2012; Pande et al. 2013; Sharma, 2021). It indicates the level of ease or difficulty of the test (Sim & Rasiyah, 2006; Guraya et al. 2018). It is determined by analyzing the scores of respondents on the individual test item (Padilla and Zabala 2017).

Based on the test scores of the respondents, the DI of each item was calculated. Each correct item was assigned one mark and after assigning the marks to all the respondents, their total score were arranged in descending order. The top one-third of the respondents were classified as high achievers (H), while the bottom one-third were designated as low achievers (L). The difficulty index (DI) was then calculated using the following formula (Mitra, Nagaraja, Ponnudurai, & Judson, 2009):

$$DI = \frac{H + L}{n} \times 100$$

where DI represents the difficulty index (expressed as a percentage), H denotes the number of high achievers, L refers to the number of low achievers, and n indicates the total number of respondents. The interpretation of calculated values of *DI* was made considering the criteria as suggested by Rao, Prasad, Permi, and Shetty (2016).

- difficulty index* < 30%too difficult
- difficulty index* = 30% - 70%appropriate
- difficulty index* > 70%too easy

The Discrimination Index

The discrimination index (*Dis I*) reflects how well an item differentiates between high and low scorers. Its value ranges from -1 to +1 (Sharma, 2021). Gajjar, Sharma, Kumar and Rana (2014) define the discrimination index as “the ability of an item to differentiate between students of higher and lower abilities” (p.18). The discrimination index, also known as the point-biserial correlation, indicates how well a test item differentiates between high and low achievers—high achievers tend to perform better on the item than low achievers (Kheyami, Jaradat, Al-Shibani, & Ali, 2018). To determine the DI for each test item, the following formula was applied, drawing on the scores obtained by the high achievers (H) and the low achievers (L) (Mitra et al. 2009; Rao et al. 2016):

$$Dis I = \frac{H - L}{n} \times 2$$

Where *Dis I* indicate the discrimination index and the results were interpreted based on the following criteria:

- Dis I* = negative (wrong item)
- Dis I* = 0 – 0.19 (poor items)
- Dis I* = 0.2 - 0.29 (acceptable)
- Dis I* = 0.3 – 0.39 (good items)
- Dis I* ≥ 0.4 (excellent items)

The Distractor Efficiency

The distractor efficiency (*DE*) reflects whether or not the test item was well constructed. It is used to judge the credibility of distractors (Sim, & Rasiah, 2006; Tarrant, Ware, & Mohammed, 2009). In the view of Mehta and Mokhasi (2014), the distractors are the most important components of a test item and significantly affect the total test scores. The performance of the examinees depends upon the designed distractors (Dufresne, Leonard, & Gerace, 2002).

In addition, Sabri (2013) and Alhazmi et al, (2024) mentioned that the quality of MCQs is associated with distractors. Assessing the effectiveness of each distractor is essential for a high-quality multiple-choice test. The effectiveness of an item's incorrect options can be determined only through DE (Denny et al. 2008). According to Ali, Carr, and Ruit (2016), a test item is more effective if the examinees with good knowledge can choose the correct option easily leaving behind the incorrect alternatives. On the other hand, the distractors appear to be the most appealing choices for examinees who possess inadequate knowledge in that area. The DE for every test item was estimated using a nonfunctional distractor. If an incorrect option is chosen by 5% or fewer respondents, it is considered a non-functional distractor. Each item on this scale consisted of four options: one correct answer and three incorrect alternatives. The DE of the item, in the case of 1 non-functional distractor will be $\left[\frac{3-1}{3}\right] \times 100 = 66.67\%$; and for 2 non-functional distractors, DE will be $\left[\frac{3-2}{3}\right] \times 100 = 33.33\%$, and for all 3 non-functional distractors the value of DE will be $\left[\frac{3-3}{3}\right] \times 100 = 0.0\%$ (Hingorjo & Jaleel, 2012).

Test Reliability

For assessing the reliability of the scale, the Kuder–Richardson formula 20 or KR-20 is commonly used which was developed by G. Frederic Kuder and M. W. Richardson (Kline, 2008; Wood et al. 2016; Dennis, Brown, & Brennenstuhl, 2018). It is employed to calculate the internal consistency of multiple-choice test items by comparing each item with the others (Alquraan, 2016). It assesses the internal consistency of multiple-choice test items by splitting the test into two equal halves. This formula is applied when each test item has a binary outcome—correct or incorrect—with one mark awarded for a correct answer and zero for an incorrect one (Beeker & Maunsaiyat, 2002). To check the reliability of the developed tool, the following Kuder-Richardson Formula 20 is used:

$$\rho_{KR-20} = \left(\frac{k}{k-1}\right) \left(1 - \frac{\sum pq}{\sigma^2}\right)$$

Where ρ_{KR-20} represents the Kuder–Richardson formula 20 reliability coefficient, k is the total number of test items; σ^2 denotes the variance of the total test scores; p is the proportion of test-takers who answered an item correctly; q is the proportion of test-takers who answered it incorrectly; and $\sum pq$ refers the sum of the products of p and q across all test items. The values of the reliability range from 0 to 1. Low values correspond to low reliability, whereas high values correspond to high test reliability (Mukherjee et al. 2019; Lin, & Liu, 2025).

In the forthcoming section, calculated values of the DI, Dis I, DE and the internal consistency reliability of the final multiple-choice test have been described and interpreted.

Results and Findings

Validation of the test was carried out by an expert panel from the field of Teacher Education. Based on their feedback, some items were revised and some were removed, reducing the total number of items from 44 to 39. After analyzing the initial draft of the MCQs test consisting of 39 items, it was found that there was 1 too difficult item, 23 were appropriate,



and 15 too easy items. Moreover, the Dis I of these 39 items reflected that there were 08 items categorized as poor items, 06 as acceptable, 06 as good, and 19 items were considered excellent.

A total of 200 secondary school science teachers participated in the multiple-choice test, which comprised 31 items with 93 distractors. Among these teachers, 66 (33%) were classified as high achievers, and another 66 (33%) were classified as low achievers. Further analyses were conducted on the test scores of these 66 high-achieving and 66 low-achieving teachers. The computed values of Dis I, DI, and DE are provided in Tables 1 and 3. Table 1 details the metrics for individual test items, while Table 3 summarizes the overall results.

Table 1:

Values of the Dis I, DI, and DE of the tryout scale.

Item No.	DI	Dis I	DE %	Item No.	DI	Dis I	DE %
1	0.80	0.27	100	21	0.51	0.12*	100
2	0.71	0.45	66.67	22	0.68	0.41	66.67
3	0.68	0.06*	100	23	0.61	0.36	100
4	0.80	0.19*	100	24	0.47	0.42	100
5	0.78	0.44	100	25	0.49	0.18*	100
6	0.85	0.33	100	26	0.82	0.32	100
7	0.83	0.24	100	27	0.82	0.30	100
8	0.54	0.58	100	28	0.38	0.53	100
9	0.58	0.14*	100	28	0.50	0.33	100
10	0.43	0.59	66.67	30	0.37	0.38	66.67
11	0.49	0.24	100	31	0.75	0.47	100
12	0.28	0.03*	100	32	0.89	0.29	66.67
13	0.59	0.44	100	33	0.52	0.17*	66.67
14	0.80	0.45	100	34	0.78	0.48	100
15	0.52	0.65	100	35	0.66	0.53	66.67
16	0.56	0.61	100	36	0.59	0.41	100
17	0.75	0.47	100	37	0.58	0.61	100
18	0.75	0.15*	100	38	0.60	0.33	100
19	0.46	0.62	100	39	0.77	0.41	100
20	0.53	0.27	100				

The test items at serial # 3, 4, 9, 12, 18, 21, 25, and 33 have the values of the Dis I ranging from 0.00 to 0.19 categorized as poor items and hence deleted. Therefore, 8 poor items were excluded from the MCQs test finalized, comprising 31 items, for administration to verify its validity and reliability. Finally, the MCQs test (Appendix A) comprised of 31 items was approved as appropriate. The detail of factors and corresponding items of the final MCQs test are shown in table 2.

Table 2:

Detail of factors and corresponding items of the final MCQs test

Factors	Number of Items	Item Labels
Orientation about Instructional Strategies	06	1, 2, 5, 6, 7, 8
Lecture/Discussion Strategies	05	10, 11, 13, 14, 15
Problem Solving Strategies	04	16, 17, 19, 20

Demonstration Strategies	04	22, 23, 24, 26
Questioning/Inquiry Strategies	06	27, 28, 29, 30, 31, 32
Cooperative Learning Strategies	06	34, 35, 36, 37, 38, 39

Table 3: Classification of test items of final scale based on the Dis I, DI, and DE.

DI	Interpretation	Mean DI	Mean DE	Items (n = 31)
> 70 %	Too easy	0.80	94.87	13 (42%)
(30 - 70) %	Appropriate	0.53	92.59	18 (58%)
< 30 %	Too difficult	0.00	0.00	0 (0%)

Dis I	Interpretation	Mean Dis I	Mean DE	Items (n = 31)
≤ 0.19	Poor	0.00	0.00	0 (0%)
0.20 - 0.29	Acceptable	0.26	93.33	5 (16%)
0.30 - 0.39	Good	0.34	95.24	7 (23%)
≥ 0.40	Excellent	0.50	92.98	19 (61%)

DI	Dis I ≤ 0.19	0.20 - 0.29	0.30 - 0.39	≥ 0.40
> 70	0	3	3	7
30 - 70	0	2	4	12
< 30	0	0	0	0
Total	0	5	7	19

The developed MCQs test comprised of 31 items, it was found that no item (0 %) was classified as too difficult (DI < 30 %) based on the estimated value of the DI. Moreover, 18 items (58 %) were classified as appropriate (30 % ≤ DI ≤ 70 %) and other 13 items (42 %) were too easy (DI > 70 %). The pictorial view of the proportions of too difficult, appropriate items and too easy items are shown in figure 1.

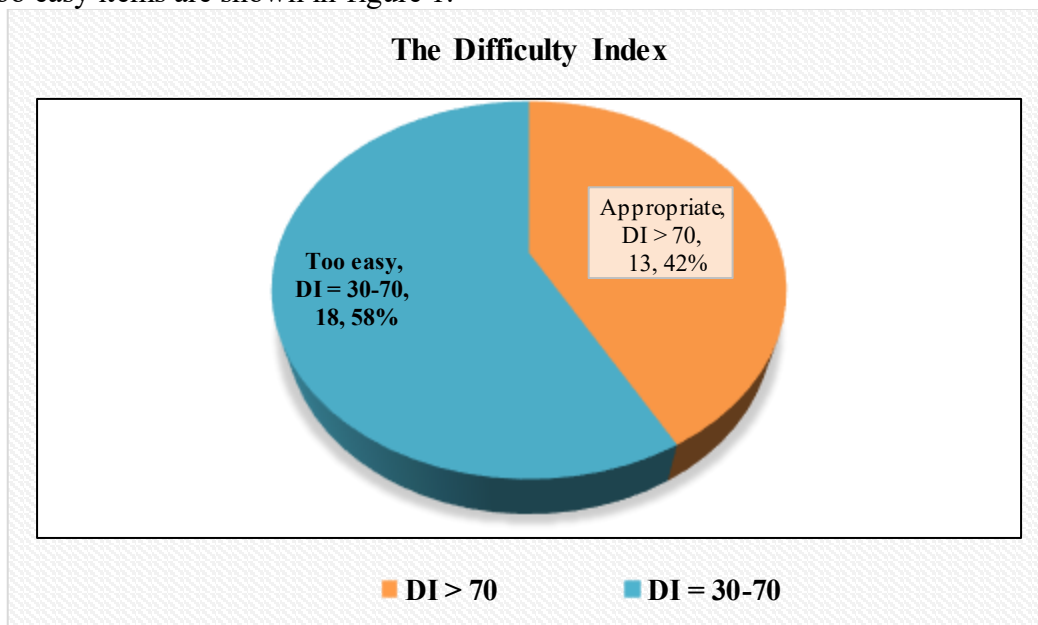


Figure 1: The DI of test items

The values of Dis I for all 31 test items were computed. In the finalized MCQs test, the Dis I values ranged from 0.24 to 0.65. None of the items fell into the poor category (Dis I = 0–0.19). A total of 5 items were classified as acceptable (Dis I = 0.20–0.29), 7 items as good (Dis I = 0.30–0.39), and 19 items as excellent (Dis I ≥ 0.40). Furthermore, no item was identified as a wrong item, as none produced a negative discrimination index. A pictorial

representation of the proportions of poor, acceptable, good, and excellent items is shown in Figure 2.

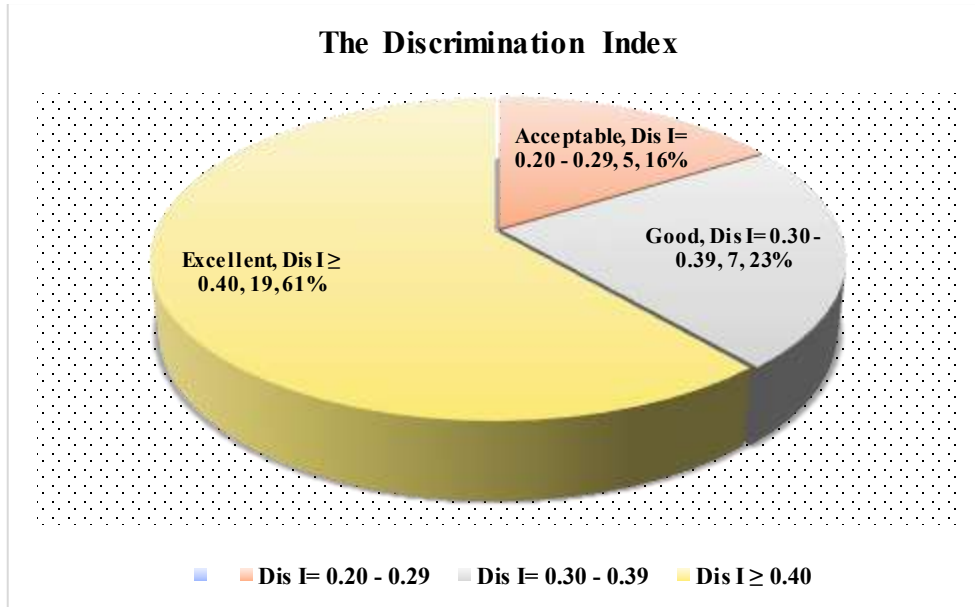


Figure 2: The Dis I of test items

In this developed tool, the distractor efficiencies of 25 items were 100% and for six items, it was found 66.67%. The values of DE of all the test items are also shown in table 1. The mean values of distractor efficiencies of the items under three categories of DI and four categories of Dis I are also shown in table 1. The mean DE of the “too easy” items with $DI > 70\%$ is 94.87% while the mean DE of the “Appropriate” items with $(30\% \leq DI \leq 70\%)$ is 92.59%. Similarly, the mean DE of the “acceptable” items with $(Dis I = 0.2 - 0.29)$ is 93.33% while the mean DE of the “good” items with $(Dis I = 0.3 - 0.39)$ is 95.24% and for “excellent” items with $(Dis I \geq 0.4)$ is 92.98%.

The internal consistency of the test items was measured with the help of Kuder-Richardson Formula 20. Measured values of p , q , and pq for each item are mentioned in table 4. The internal consistency value (ρ_{KR-20}) of the developed MCQs test was 0.85 which is quite appropriate level for tests with item counts of 50 or less (Fu et al. 2019; Govindasamy et al, 2024). Moreover, test with reliability value of more than 0.9 is considered a standardized test (Yung, 2009). Other calculated values were: $K = 31$; $\Sigma pq = 6.46$; $\rho_{KR20} = 0.85$.

Table 4

The item-wise values according to the Kuder–Richardson Formula 20.

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
p	0.80	0.71	0.78	0.85	0.83	0.54	0.43	0.49	0.59	0.80	0.52	0.56	0.75	0.46	0.53	0.68
q	0.20	0.30	0.23	0.15	0.17	0.47	0.57	0.52	0.41	0.20	0.48	0.45	0.25	0.54	0.47	0.33
pq	0.16	0.21	0.17	0.13	0.14	0.25	0.25	0.25	0.24	0.16	0.25	0.25	0.19	0.25	0.25	0.22
Item	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	
p	0.61	0.47	0.82	0.82	0.38	0.50	0.37	0.75	0.89	0.78	0.66	0.59	0.58	0.60	0.77	
q	0.40	0.53	0.19	0.19	0.63	0.50	0.63	0.25	0.12	0.23	0.35	0.42	0.42	0.40	0.23	
pq	0.24	0.25	0.15	0.15	0.23	0.25	0.23	0.19	0.10	0.17	0.23	0.24	0.24	0.24	0.18	

Discussion and Conclusion

The pictorial view of the values of DI and $Dis I$ for all test items is presented in figure 3. The values of DI and $Dis I$ reflects that there are neither “too difficult” nor “poor” test

items in the MCQs test. However, there are two items at serial numbers 21 and 23 with *DI* values of 0.38 and 0.37 that are “appropriate” but close to the criteria ($DI < 30\%$) of the “too difficult” level. On the other hand, these two items with *Dis I* values of 0.53 and 0.38 under the criteria of *Dis I* are “Excellent” and “Good” respectively. Moreover, two items at serial numbers 5 and 8 with common *Dis I* value of 0.24 are on the “acceptable” and “too easy” levels under the criteria of *DI* with values of 0.83 and 0.49 respectively. In the nutshell, results indicate that all 31 test items are ideal for measuring science teachers’ knowledge about instructional strategies.

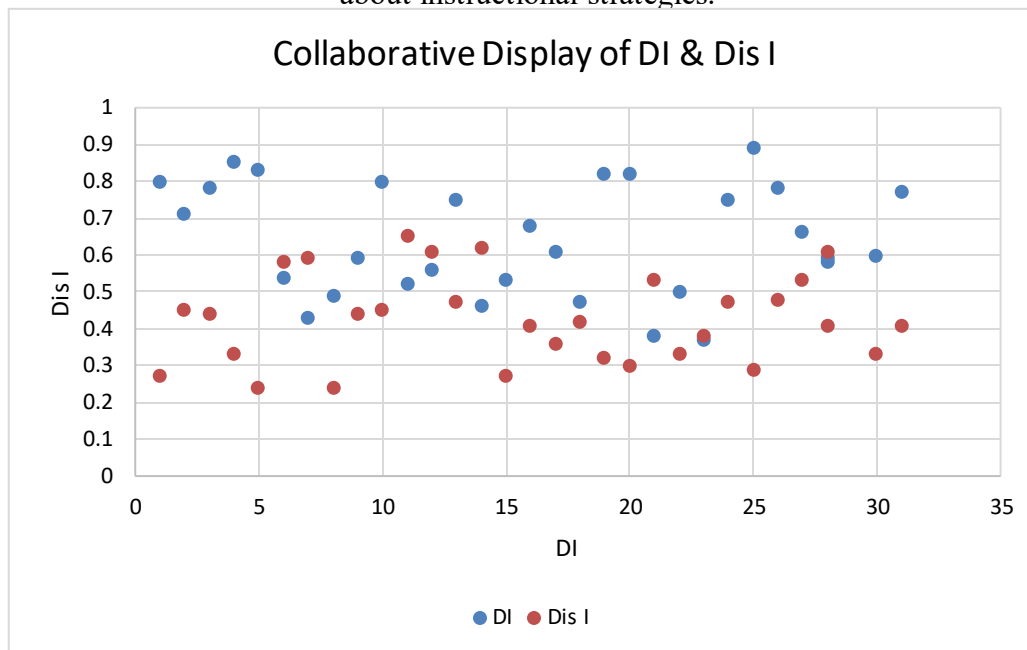


Figure 3: Collaborative Display of *DI* & *Dis I*

To assess the internal consistency reliability in the present study, the Kuder–Richardson Formula 20 was applied, yielding a coefficient value of 0.85. Several studies indicate that this value of internal consistency represents an acceptable level of reliability. For instance, the value of the reliability coefficient of the MCQs test developed by Iqbal, Saleem, and Arshad, (2020) was 0.75. Moreover, the Assessment Literacy Inventory (ALI), developed by Mertler and Campbell (2005), showed an internal consistency reliability value of 0.74, indicating an acceptable level of reliability. Furthermore, some other developed MCQs tests have internal consistency reliability through KR20 ranging from 0.74 to 0.80 (Iqbal, Saleem, & Arshad, 2020). Thus, the findings indicate that the MCQs test constructed for this study possesses satisfactory internal consistency reliability.

The above-mentioned results indicate that all of the test items of the developed MCQs test are appropriate for measuring different aspects of science teachers’ knowledge about instructional strategies i.e., orientation about instructional strategies, knowledge of lecture/discussion strategies, knowledge of problem-solving strategies, knowledge of demonstration strategies, knowledge of questioning/inquiry strategies, and knowledge of cooperative learning strategies.

Future Research Directions

The MCQs test developed in this study serves as a useful tool for assessing secondary school science teachers’ knowledge of instructional strategies. Its structured format and carefully constructed items make it suitable for evaluating how well teachers understand and



apply various pedagogical approaches. To further strengthen the generalizability and robustness of the test, its reliability and validity may be examined by administering it to teachers at different educational levels, such as primary, higher secondary, and college sectors. This broader application would help determine whether the test performs consistently across diverse teaching contexts.

In future research, the psychometric properties of the test can be enhanced by revising the response format—for example, converting the MCQs into Likert-scale items to capture more nuanced responses. Additionally, employing advanced statistical techniques, such as Cronbach's alpha for internal consistency and exploratory factor analysis for construct validation, could provide deeper insights into the test's measurement quality. These methods would serve as complementary approaches to the traditional item analysis procedures used in this study, such as the Dis I, DI, DE, and the Kuder–Richardson Formula 20. Together, these extensions would contribute to a more comprehensive evaluation of the test's effectiveness and reliability.

As for as the scoring of the MCQs test is concerned, each correct item is marked as a “1” score and each incorrect answer is marked as a “0” score. The scale for assessing the levels of knowledge is as under:

- Test score: Below 50%Poor
- Test score: 50% - 75%Good
- Test score: Above 75%Excellent

Funding

This manuscript was not funded.

Disclosure statement

The authors have no conflicts of interest to disclose.

References

- Alhazmi, E., Sheng, Q., Zhang, W. E., Zaib, M., & Alhazmi, A. (2024). Distractor generation in multiple-choice tasks: A survey of methods, datasets, and evaluation. In *Proceedings of the 2024 conference on empirical methods in natural language processing* (pp. 14437-14458).
- Ali, S. H., Carr, P. A., & Ruit, K. G. (2016). Validity and Reliability of Scores Obtained on Multiple-Choice Questions: Why Functioning Distractors Matter. *Journal of the Scholarship of Teaching and Learning*, 16(1), 1-14.
- Alquraan, M. (2016). DIF items effects on Kuder-Richardson 20 formula and construct validity: an exploration using the Mantel-Haenszel method. *International Journal of Quantitative Research in Education*, 3(4), 266-276.
- Becker, K. H., & Maunsaiyat, S. (2002). Thai Students' Attitudes and Concepts of Technology. *Journal of Technology Education*, 13(2), 6-20.
- Beullens, J., Damme, B. V., Jaspert, H., & Janssen, P. J. (2002). Are extended-matching multiple-choice items appropriate for a final test in medical education? *Medical Teacher*, 24(4), 390-395.
- Christian, D. S., Prajapati, A. C., Rana, B. M., & Dave, V. R. (2017). Evaluation of multiple-choice questions using item analysis tool: a study from a medical institute of Ahmedabad, Gujarat. *International Journal of Community Medicine And Public Health*, 4(6), 1876.
- DiBattista, D., & Kurzawa, L. (2011). Examination of the quality of multiple-choice items on classroom tests. *Canadian Journal for the Scholarship of Teaching and Learning*, 2(2), 4.



- Darling-Hammond, L. (2000). Teacher quality and student achievement. *Education policy analysis archives*, 8, 1.
- Dennis, C. L., Brown, H. K., & Brennenstuhl, S. (2018). Development, Psychometric Assessment, and Predictive Validity of the Postpartum Childcare Stress Checklist. *Nursing Research*, 67(6), 439-446.
- Denny, P., Hamer, J., Luxton-Reilly, A., & Purchase, H. (2008). PeerWise: students sharing their multiple-choice questions. In *Proceedings of the fourth international workshop on computing education research* (pp. 51-58).
- Depaepe, F., Verschaffel, L., & Kelchtermans, G. (2013). Pedagogical content knowledge: A systematic review of the way in which the concept has pervaded mathematics educational research. *Teaching and Teacher Education*, 34, 12–25.
- Dufresne, R. J., Leonard, W. J., & Gerace, W. J. (2002). Marking sense of students' answers to multiple-choice questions. *The Physics Teacher*, 40(3), 174-180.
- Earnest, B. S. P., Bhargava, P., Ponnupillai, A., Ponnusamy, S., Ibrahim, N. M., & Sirisinghe, R. G. (2018). Appropriate Use of Technology: How Useful are Calculations of Discrimination Index by Optical Mark Readers in Item Analysis of Single Best Answer MCQ Tests with Small Student Numbers? In *Redesigning Learning for Greater Social Impact* (pp. 267-275). Springer, Singapore.
- Fu, E., White, M. A., Hughto, J. M., Steiner, B., & Willis, E. A. (2019). Development of the Physical Activity Tracking Preference Questionnaire. *International journal of exercise science*, 12(5), 297.
- Gajjar, S., Sharma, R., Kumar, P., & Rana, M. (2014). Item and test analysis to identify quality multiple-choice questions (MCQs) from an assessment of medical students of Ahmedabad, Gujarat. *Indian Journal of Community Medicine*, 39(1), 17-22.
- Gierl, M. J., Bulut, O., Guo, Q., & Zhang, X. (2017). Developing, analyzing, and using distractors for multiple-choice tests in education: a comprehensive review. *Review of Educational Research*, 87(6), 1082-1116.
- Govindasamy, P., Cumming, T. M., & Abdullah, N. (2024). Validity and reliability of a needs analysis questionnaire for the development of a creativity module. *Journal of Research in Special Educational Needs*, 24(3), 637-652.
- Gözütok, F. D. (2000). *Öğretmenliği Geliştiriyorum*, Ankara: Siyasal Yayıncılık.
- Großschedl, J., Harms, U., Kleickmann, T., & Glowinski, I. (2015). Preservice biology teachers' professional knowledge: Structure and learning opportunities. *Journal of Science Teacher Education*, 26(3), 291–318.
- Großschedl, J., Welter, V., & Harms, U. (2019). A new instrument for measuring pre-service biology teachers' pedagogical content knowledge: The PCK-IBI. *Journal of Research in Science Teaching*, 56(4), 402-439.
- Guerriero, S. (2014). Teachers' pedagogical knowledge and the teaching profession. *Teaching and Teacher Education*, 2(1), 7.
- Guerriero, S. (ed.) (2017), *Pedagogical Knowledge and the Changing Nature of the Teaching Profession*, Educational Research and Innovation, OECD Publishing, Paris, <https://dx.doi.org/10.1787/9789264270695-en>.
- Guraya, S. Y., S. S. Guraya, F. Habib, K. W. AlQuiliti, and K. I. Khoshhal. (2018). Medical Students' Perception of Test Anxiety Triggered by Different Assessment Modalities. *Medical Teacher* 40 (sup1): S49–S55.
- Haladyna, T. (2013). *Selected-Response Format: Developing Multiple-Choice Items*. 1st ed. chapter 5; New York, NY, Routledge, 79-131.



- Hingorjo, M. R., & Jaleel, F. (2012). Analysis of one-best MCQs: the difficulty index, discrimination index and distractor efficiency. *JPMA-Journal of the Pakistan Medical Association*, 62(2), 142.
- Iqbal, Z., Saleem, K., & Arshad, H. M. (2020). Measuring teachers' knowledge of student assessment: Development and validation of an MCQ test. *Educational Studies*, 1-18.
- Kheyami, D., Jaradat, A., Al-Shibani, T., & Ali, F. A. (2018). Item analysis of multiple-choice questions at the department of paediatrics, Arabian Gulf University, Manama, Bahrain. *Sultan Qaboos University Medical Journal*, 18(1), e68-e74.
- Kimani, G. N., Kara, A. M., & Njagi, L. W. (2013). Teacher factors influencing students' academic achievement in secondary schools in Nyandarua County, Kenya. *International journal of education and research*, 1(3), 1-14.
- Kıyak, Y. S., Budakoğlu, I. İ., Bakan Kalaycıoğlu, D., Kula, S., and Coşkun, Ö. (2022). Can preclinical students improve their clinical reasoning skills only by taking case-based online testlets? A randomized controlled study. *Innov. Educ. Teach. Int.* 60, 1–10. doi: 10.1080/14703297.2022.2041458.
- Kleickmann, T., Großschedl, J., Harms, U., Heinze, A., Herzog, S., Hohenstein, F., ... Zimmermann, F. (2014). Professionswissen angehender Lehrkräfte mit mathematisch-naturwissenschaftlichen Fächern: Testentwicklung im Rahmen des Projekts KiL [Professional knowledge of preservice mathematics and science teachers: Test construction in the project KiL]. *Unterrichtswissenschaft*, 42(3), 280–288.
- Kline, R. B. (2008). *Becoming a behavioural science researcher: A guide to producing research that matters*. Guilford Press.
- Kuechler, W. L., & Simkin, M. G. (2010). Why is performance on multiple-choice tests and constructed-response tests not more closely related? Theory and an empirical test. *Decision Sciences Journal of Innovative Education*, 8(1), 55-73.
- Lee, E., & Luft, J. A. (2008). Experienced secondary science teachers' representation of pedagogical content knowledge. *International Journal of Science Education*, 30, 1343–1363.
- Lin, C. H., & Liu, M. Y. (2025). Development and Validation of the Eating Support for Healthcare Aides (ESHA) Questionnaire in Long-Term Care. *Nutrients*, 17(20), 3235.
- Lohse-Bossenz, H., Kunina-Habenicht, O., & Kunter, M. (2013). The role of educational psychology in teacher education: expert opinions on what teachers should know about learning, development, and assessment. *European Journal of Psychology of Education*, 28(4), 1543-1565.
- Malau-Aduli, B. S., & C. Zimitat. (2012). Peer Review Improves the Quality of MCQ Examinations. *Assessment & Evaluation in Higher Education* 37(8): 919–931.
- Marzano, R. J. (2007). *The art and science of teaching: A comprehensive framework for effective instruction*. Alexandria, VA: Association for Supervision and Curriculum Development. *British Journal of Educational Psychology*, 75(3), 351-372.
- Marzano, R. J. (2017). *The new art and science of teaching*. Bloomington, IN: Solution Tree Press.
- Mehta, G., & Mokhasi, V. (2014). Item analysis of multiple-choice questions: An assessment of the assessment tool. *International Journal Health Science Research*, 4(7),197-202.
- Mertler, C. A., & Campbell, C. (2005). Measuring Teachers' Knowledge & Application of Classroom Assessment Concepts: Development of the "Assessment Literacy Inventory". Paper presented at the annual meeting of the American Educational Research Association, Montreal, Quebec, Canada, April 11–15.



- Meyen, E., & Greer, D. (2009). The role of instructional planning in math instruction for students with learning disabilities. *Focus on exceptional children*, 41(5), 1.
- Mitra, N. K., Nagaraja, H. S., Ponnudurai, G., & Judson, J. P. (2009). The levels of difficulty and discrimination indices in type A multiple-choice questions of pre-clinical semester 1 multidisciplinary summative tests. *IeJSME*, 3(1), 2-7.
- Mukherjee, A., Singh, P., Burman, R. R., Shubha, K., & Roy, M. L. (2018). Development of a test to measure knowledge level of farmers producer organization members about improved hill agricultural practices. *Journal of Community Mobilization Sustainable Development* 14(1), 57–64.
- Padilla, E. B., & Zabala, J. L. T. (2017). Tested!: From Assessment Tool Test to Session Revision of an Information Literacy Instruction Module. *PAARL*, 40-57.
- Palmer, E. J., & Devitt, P. G. (2007). Assessment of higher-order cognitive skills in undergraduate education: modified essay or multiple choice questions? Research paper. *BMC medical education*, 7(1), 1-7.
- Pande, S. S., Pande, S. R., Parate, V. R., Nikam, A. P., & Agrekar, S. H. (2013). Correlation between difficulty and discrimination indices of MCQs in the formative exam in physiology. *South-East Asian Journal of Medical Education*, 7(1), 45-50.
- Rao, C., Prasad, H. K., Sajitha, K., Permi, H., & Shetty, J. (2016). Item analysis of multiple-choice questions: Assessing an assessment tool in medical students. *International Journal of Educational and Psychological Researches*, 2(4), 201.
- Ruys, I., Keer, H. V., & Aelterman, A. (2012). Examining pre-service teacher competence in lesson planning pertaining to collaborative learning. *Journal of Curriculum Studies*, 44(3), 349-379.
- Sabri, S. (2013). Item analysis of student comprehensive test for research in teaching beginner string ensemble using model-based teaching among music students in public universities. *International Journal of Education and Research*, 1(12), 1-14.
- Santiago, P. and F. Benavides (2009), Teacher Evaluation: A Conceptual Framework and Examples of Country Practices, paper presented at the OECD Evaluation Framework in Mexico: International Prac City, 1-2 December 2009, OECD, Paris, www.oecd.org/edu/evaluationpolicy.
- Sharma, L. R. (2021). Analysis of Difficulty Index, Discrimination Index and Distractor Efficiency of Multiple-Choice Questions of Speech Sounds of English. *International Research Journal of MMC*, 2(1), 15-28.
- Sim, S. M., & Rasiyah, R. I. (2006). Relationship between item difficulty and discrimination indices in true/false-type multiple choice questions of a para-clinical multidisciplinary paper. *Annals-Academy of Medicine Singapore*, 35(2), 67-71.
- Sutopo, A., Rahman, A., & Mulyana, D. (2024). Design and Validation of a Competency Assessment for Electrical Power Systems in Engineering Education. *AL-ISHLAH: Jurnal Pendidikan*, 16(4), 4588-4598.
- Tarrant, M., Ware, J., & Mohammed, A. M. (2009). An assessment of functioning and non-functioning distractors in multiple-choice questions: a descriptive analysis. *BMC medical education*, 9(1), 1-8.
- Turkmen, H. (2009). Examining elementary science education teacher's disposition after reform. *Asia-Pacific Forum on Science Learning and Teaching*, 10(2), 1-9.
- Voss, T., Kunter, M., & Baumert, J. (2011). Assessing teacher candidates' general pedagogical/psychological knowledge: Test construction and validation. *Journal of educational psychology*, 103(4), 952.



- Watson, M. (2003). *Learning to trust: Transforming difficult elementary classrooms through developmental discipline*. San Francisco: Jossey-Bass.
- Wood, A., Holland, S., Mozer, M., Tangney, C., & Rasmussen, H. (2016). Development and Psychometric Testing of a Questionnaire to Assess Knowledge of and Stages of Change Related to a Low-Sodium Diet in Heart Failure Patients. *The FASEB Journal*, 30, 1b334-1b334.
- Yung, H. I. (2009). Effects of an Animated Pedagogical Agent with Instructional Strategies in Multimedia Learning. *Journal of educational multimedia and hypermedia*, 18(4), 453.

Appendix A: MCQs Test to Assess the Science Teachers' Knowledge about Instructional Strategies

Name of Teacher (Optional): Gender: Male/Female:
Academic Qualification: Professional Qualification:
School location: Urban/Rural: Teaching Experience: (Years)
Choose the correct option.

Orientation about Instructional Strategies

1. Lesson plan makes the instruction:
 - a. Regular
 - b. Systematic
 - c. Organized
 - d. All of the above
2. The first step in lesson planning is to:
 - a. Select material
 - b. Set objectives
 - c. Brainstorming
 - d. Specify content
3. The basic requirement of effective teaching:
 - a. Expertise in the subject matter
 - b. Mastery of teaching skills
 - c. Mastery of appropriate use of teaching resources
 - d. All of the above
4. An effective teacher should:
 - a. Provides a variety of learning experiences
 - b. Engage students in book reading
 - c. Deliver long lectures
 - d. Keep students busy
5. An effective teacher makes students attentive in class:
 - a. By distributing sweets
 - b. By putting naughty children to task
 - c. By making one's teaching interesting
 - d. By telling stories
6. The general purpose of home assignments is:
 - a. To foster the habits of consistent independent study
 - b. To provide an opportunity for student learning in the classroom
 - c. To improve students' social skills
 - d. To keep students busy

Lecture/Discussion Strategies



7. A teacher's role in the lecture method is like a:
 - a. Partner
 - b. Facilitator
 - c. Delegator
 - d. Formal authority
8. The most suitable method to convey a lot of information about a topic is:
 - a. Lecture method
 - b. Inquiry-based method
 - c. Cooperative method
 - d. Assignment method
9. A major disadvantage of the lecture method is:
 - a. Dependence on memory
 - b. Lack of self-thinking
 - c. Lack of scientific attitude
 - d. All of the above
10. Students are passive in
 - a. Project method
 - b. Discovery method
 - c. Lecture method
 - d. Inquiry method
11. Which of the following statements is not correct about the lecture method?
 - a. The lecture method can develop reasoning
 - b. The lecture method can develop knowledge
 - c. The lecture method is a one-way process
 - d. During the lecture method, students are passive

Problem Solving Strategies

12. To develop scientific skills among students, the most suitable teaching strategies is:
 - a. Lecturing strategy
 - b. Questioning Strategy
 - c. Cooperative learning strategy
 - d. Problem-solving strategy
13. In the problem-solving method, the first step is:
 - a. Design a solution
 - b. Define a problem
 - c. Searching the solution
 - d. Organizing the data
14. The limitation of the problem-solving method is:
 - a. Students learn to find a solution to their problem
 - b. This method follows the principle of learning by doing
 - c. Students learn to use old facts in new references
 - d. There is always doubt about drawing the wrong conclusions
15. The type of problem-solving method is:
 - a. Inductive and deductive instruction
 - b. Programmed instruction
 - c. Situational instruction
 - d. Strategic instruction

Demonstration Strategies



16. The teacher performs practically and explains in:
 - a. Lecture method
 - b. Discovery method
 - c. Demonstration method
 - d. Problem-solving method
17. To improve the students' comprehension, a teacher may:
 - a. Take the help of notes
 - b. Give illustrations to clarify concepts
 - c. Deliver long lectures
 - d. All of the above
18. At the end of the demonstration, it is appreciable to:
 - a. Evaluate demonstration
 - b. Keep the demonstration simple
 - c. Ensure students' active participation
 - d. Provide short notes
19. An effective demonstration in teaching comprises:
 - a. Audio presentation
 - b. Visual presentation
 - c. Audio-visual presentation
 - d. Graphic presentation

Questioning/Inquiry Strategies

20. For the involvement of students, a teacher may:
 - a. Deliver long lectures
 - b. Ask questions from students
 - c. Use blackboard
 - d. Tell stories to students
21. The combination of teaching methods to create a learner-centred situation is?
 - a. Individualized instruction and lecture method
 - b. Lecture and demonstration
 - c. Lecture method and experimentation
 - d. Problem-solving and questioning
22. The teacher's role in inquiry-based learning is as:
 - a. Instructor
 - b. Facilitator
 - c. Delegator
 - d. Formal authority
23. The Socratic Method of teaching is also known as:
 - a. Lecture-demonstration method
 - b. Discussion method
 - c. Inquiry method
 - d. Question-answer method
24. The students who ask questions in the class should be:
 - a. Advised to meet the teacher after the class
 - b. Encouraged to continue asking questions
 - c. Encouraged to search answers independently
 - d. Advised to consult the textbook
25. A teacher asks the questions in the class to:



- a. Keep students busy
- b. Maintain discipline
- c. Attract student's attention
- d. Make students ashamed

Cooperative Learning Strategies

26. Any kind of assignment that involves two or more students, is a form of:

- a. Collaborative learning
- b. Demonstration
- c. Problem-solving
- d. Lecture method

27. To make students socially active, they may be:

- a. Involved in group activities
- b. Involved in book-reading
- c. Encouraged to search answers independently
- d. Encouraged to continue asking questions

28. Cooperative learning is a:

- a. Teacher-centred approach
- b. Student-centred approach
- c. Subject-centred approach
- d. Problem-centred approach

29. Which is the student-centred learning method?

- a. Direct instruction
- b. Book-reading method
- c. Cooperative learning
- d. Lecture method

30. The major advantage of cooperative learning is:

- a. It creates a classroom of behaviour problems
- b. Requires adequate facilities and infrastructure
- c. The students are much excited about learning
- d. The students have no concern with each other

31. In cooperative learning, students are:

- a. Passive learner
- b. Allowed discussion and critical thinking
- c. Independent
- d. Spectators

Key to Scale

1. All of the above
2. Set objectives
3. All of the above
4. Provides a variety of learning experiences
5. By making one's teaching interesting
6. To foster the habits of consistent independent study
7. Formal authority
8. Lecture method
9. All of the above
10. Lecture method
11. The lecture method can develop reasoning



12. Problem-solving strategy
13. Define a problem
14. There is always a doubt about drawing the wrong conclusions
15. Inductive and deductive instruction
16. Demonstration method
17. Give illustrations to clarify concepts
18. Evaluate demonstration
19. Audio-visual presentation
20. Ask questions from students
21. Problem-solving and questioning
22. Facilitator
23. Question-answer method
24. Encouraged to continue asking questions
25. Attract student's attention
26. Collaborative learning
27. Involved in group activities
28. Student-centred approach
29. Cooperative learning
30. The students are much excited about learning
31. Allowed discussion and critical thinking