

# Building a Criterion-Referenced Test in Measurement and Evaluation and Determining Its Cut-Off Score by Several Methods

Hani Alkhalidi<sup>1</sup> Malek Alkhubaba<sup>2</sup> Sami Almassarweh<sup>2</sup> Hussein Hikmat Al-Mistareehi<sup>1</sup>  
Abdul-Raouf Hamid Al-Yamani<sup>3</sup>

1. Department of teacher education, Isra University, Jordan

2. Department of Psychology, Isra university, Jordan

3. Department of Child Education, Isra University, Jordan

Corresponded author Malek Alkhubaba, [malek.alkhubaba@iu.edu.jo](mailto:malek.alkhubaba@iu.edu.jo)

## Abstract

The current study aimed to build a criterion-referenced test in measurement and evaluation and determine its cut-off score by several methods. The primary test form had 45 items, which a group of professors and measurement and evaluation experts reviewed; their comments and feedback were taken into account, and the final test form had 40 items. The test has been presented to 174 university students to examine its psychometric characteristics. Multiple statistical techniques were later performed using the SPSS program, and the results show that the discrimination and difficulty coefficients ranged from 0.36 to 0.82. Additionally, the test reliability was calculated using the Kuder-Richardson -20 and Spilt half statistical methods, and the concurrent validity was 0.76. The results showed that the value of the Kuder-Richardson -20 method was 0.81, while the value of the Spilt-Half method was 0.79. Finally, the cut-off score has been calculated using four methods, and the results indicate that the Angoff method value was 65%, the Nedelsky method was 64%, the contrasting groups' method was 68%, and the criterion groups' method was 62%.

**Keywords:** criterion-referenced test, measurement, and evaluation, university student, the cut-off score.

**DOI:** 10.7176/JEP/14-1-06

**Publication date:** January 31<sup>st</sup> 2023

## 1. Interdiction

Achievement tests are considered one of the main pillars of the educational process, as they provide important quantitative information and data that allow the test holder to judge the extent to which educational objectives have been achieved at various levels, and it also helps the student to know his scientific level through the mark of the test (Abdel-Salam, 1992). Modern measurement faced sharp criticisms of the traditional theory of psychological and educational measurement, including the weak independence of measurement results if applied to two different tests, which called for the emergence of contemporary attitudes in psychological and educational measurement to reach accuracy, and objectivity in measurement. So, the relationship between the measurement tool and the underlying trait of the examinee is very strong and accurate (Nunnly, 1978).

Psychometrics is described as a relative measurement, the marks achieved by the student on the test have no meaning except by returning it to an appropriate standard that can be relied upon to interpret it, so there are two attitudes to interpret the results of achievement tests that differ from each other: first, Norm-referenced in which the interpretation of the examinee score is based on its comparison with the average achievement of the test standard (Sigmon and Haplin, 1983). Second, the criterion-referenced test which performance is interpreted by the examinee based on the number of learning objectives it has achieved, helps us identify failures in the learning and teaching process (Allam, 2001; Allam, 2000). The criterion-referenced test can be used in a mastery learning evaluation,

The acceptable level of performance can be determined in the form of a numerical estimate, such as not being less than 90% to be classified as a proficient student, and criterion-referenced be used to make educational decisions (Zamili, 2021). Furthermore, it can be used to make diagnostic decisions and evaluate achievement in outcome-based education that arose within the framework of interest in total quality management. The goal of evaluation has changed from making judgments about how much students know of the content, and the time required to teach that content, to set standards for what students are expected to know and do (Ababneh, 2009). Ebel (1979) pointed out that the criticisms of standardized reference tests, lead to the emergence of new developments in the methodology of the tests which are called criterion-referenced tests.

These tests aim to determine what a student can and cannot master and judge based on a predetermined criterion called a cut-off score. Al-Atwi and Al-Masoudi's pointed to the low level of student performance in the subject of measurement and evaluation, and because the two specializations of measurement and evaluation and my colleagues are clinical psychology, so it was necessary to ensure the extent to which students master the most important basics and skills of measurement and evaluation, and based on which students are classified into

proficient and non-proficient based on the degree of cutting that will be determined (Al-Atwi and Al-Masoudi, 2019).

Therefore, the step of determining the cut-off score is a basis for building the criterion references test, through which the basic assumptions on which the measurement is based are verified by the criterion references test, it provides the test designer with quantitative estimates to determine the extent to which students can skill, judge its performance and then make the appropriate decision, so determining the degree of cutting for the test directly affects those decisions, any error resulting from determining the degree of cutting directly affects and thus leads to negative results (Allam, 1986).

## 2. Previous studies

Zamili (2021) aimed at building a criterion-reference test in descriptive statistics subjects using the single parameters Logistic Model according to the item's response theory. The final form of the test was (40) multiple-choice items and the sample of the study consisted of (90) students. The finding of the study indicated that the assumptions of the item's response theory were fulfilled, moreover, two items that did not match the model were deleted, and (5) participants of the study sample were deleted that did not match the model. Findings also showed the test had several psychometric properties: the reliability of the test was (0.78) and the reliability of individuals' abilities was (0.80).

Olimat (2020) aimed to construct "a criterion-referenced test in mathematics using the Three Parameters Logistic Model". In order to achieve the study aims, the researcher was building a test consisting of (35) multiple-choice items with four alternatives. The sample study consisted of 968 participants. "One item with negative discrimination was deleted. In addition, 3 items did not fit the used model. The value of Livingstone's coefficient was equal to 0.83. The test empirical reliability coefficient was 0.833". Results of the study also presented that item parameters estimation comprising difficulty, guessing, and discrimination was acceptable within the indexes available in educational measurement literature. The cut-off score using the Angoff method was (0.22). In another study, Al-Atwi and Al-Masoudi (2019) aimed to build a criterion-reference test to measure the competencies of statistics among postgraduate students according to the Rush model, the researchers used the analysis descriptive approach. A criterion -reference has been built consisting of (47) items covering red the necessary competencies for statistics, and the difficulty coefficient, validity, and reliability of the test have been examined. The findings of the study indicated the matching of test items and sample data to the Rush model, in addition to determining the cut-off score Angoff method which was ranging between 0.32 to 0.47.

Fidelis and Tei-Firstman (2017) "identified if there is a significant difference between using the Angoff method and the norm-referenced method in the setting of cut-off scores in the school setting". The study made use of 80 (JSS 3) Basic 9 students from Nembe Loca Government Area of Bayelsa state. The sample was drawn through a simple random sampling method. The design of the study was a comparative analysis. A forty-item multiple choice objective test on mathematics which was tested for goodness of fit using the big step software was used. "The internal consistency which was determined by Cronbach alpha was 0.64 while two research questions directed the conduct of the work". Percentages and intra-class correlation coefficient (ICC) were used to analyze the data collected. The comparative analysis made between the setting of the cut-off score using the Angoff and norm-reference method has a significant difference.

## 3. Statement of Problem

The educational process in faculties of educational sciences requires determining and evaluating the levels of knowledge and students' academic skills. In order to help the students to reach levels of thinking, understanding, and proficiency, educational experiences are provided, to know their tendencies and preparations, and attitudes toward various issues. This requires building accurate achievement tests that enjoy a high level of validity and reliability to evaluate the differences between the skills and knowledge of the students. However, there is a clear deficiency in the building of achievement tests, especially criterion reference tests, which are the best and most effective methods to examine and investigate the level of student performance, where the cut-off mark is still fixed at 50% for students who pass in the university subjects.

Therefore, the cut-off score must vary according to the subject's area, as students in medicine subjects must be perfecting most of the basic medical because it is related to human life. So, it's logical to be the cut-off mark around 90%. In another hand, the basic concepts of social sciences subjects are limited and popular as culture, it's logical to be the cut-off mark around 50%. In the current study, a criterion reference test has been built in measurement and evaluation subjects which consider the most important subject among students of educational social sciences. The cut-off score for the measurement and evaluation subjects is expected to be between 60% to 70%.

Thus, the professionals in measurement and evaluation have differed about the appropriate method to determine and fix the suitable cut-off marks. The bases steps have been presented to determine the suitable cut-off mark that can be trusted in the evaluation and classification of students into standardized and unstandardized

groups, with trying to avoid the error of overlap between the two categories, which often results from measurement errors e.g., coincidence factor, cheating factor, and guessing factor. The problem of the current study determined building criterion reference tests in measurement and evaluation subjects, in addition to investigating its psychometric characters and determining its cut-off score in several ways.

#### **4. Significance of the study**

The importance of the current study represented in increasing the knowledge of researchers and faculty members about the importance of the cut-off score, evaluating the proficiency of the students, and determining the test cut-off score in several ways Angoff, Nedelsky, contrasting groups, and groups' criterion methods. Furthermore, the faculty members can use the test during their teaching methods to evaluate the academic skills and proficiency of the students in measurement and evaluation subjects. In addition, diagnosing and examining students' weaknesses to design good plans for treatment and improve the students' learning skills.

#### **5. Research Questions**

The current study tries to answer the following questions:

RQ1. What are the indications of the difficulty and discrimination coefficients for the test items?

RQ2. What are the indications of the test's concurrent validity and test reliability?

RQ3. What is the cut-off score for passing the measurement and evaluation test according to Angoff, Nedelsky, contrasting groups, and criterion groups' method?

RQ4. What is the sensitive indication of the item at the cut-off score according to the Angoff method?

#### **6. Objectives of the study**

The current study aims to build a criterion-referenced test in measurement and evaluation for university students. Furthermore, to determine its cut-off score in several methods

#### **7. Terminology**

Criterion-referenced test: "the test designed to measure student performance against a fixed set of predetermined criteria or learning standards" (Gray,1978: 87).

Procedural definition: a test of basic skills and competencies in measurement and evaluation subject for the faculty of educational sciences at Isra University, where the test consists of (40) multiple-choice items.

Cut-off score: "the lowest possible score on an exam, standardized test, high-stakes test, or other forms of assessment that a student must earn to either pass or be considered proficient" (Hambleton, 1978: 121).

Procedural definition: It is the mark resulting from the application of the Angoff method to the test prepared to classify students as proficient or non-proficient.

Angoff method: a method used in an objective and multiple-choice test that requires each examiner to predict the difficulty level of each item in the test. The examiners are asked to visualize how many students have the minimum proficiency to pass, then estimate the probability that the student with the minimum proficiency will correctly know the answer to the items without resorting to guesswork. Later on, add the probability values of all items for all the examiners and the average of these probabilities is equal to the cut-off score for that test (Angof, 1971).

Nedelsky Method: This m requires each examiner to predict the difficulty level examine the items of the test, then determine from the wrong alternatives for each item those that can be avoided by the student with a minimum of adequacy. The minimum probability of the correct answer to the item is the reciprocal of the number of remaining alternatives (Meskanska, 1976).

Contrasting group: in this method the examiners must divide the students into two groups, a standardized group, and a non-standardized group. After examining the two groups, the examiners must make two curves for the student's marks distribution. Where the cross point of the two curve represented the cut-off score, in another hand the examiner can modify the first classification error 'Alpha' and the second classification error 'Beta' (Berk,1980).

Criterion groups: this method aims to limit the errors of the wrong classifications and increase the correct classifications. It's somewhat similar to the contrasting group method but requires finding binary classifications for the following probabilities: real mastered students, real non-mastered students, mastered students are non-real, and non-mastered students are non-real. Later on, determine the validity coefficient of the possible cut-off scores, and calculate the relative benefit and harm of classification errors (Berk,1982).

#### **8. Methodology**

The researchers used the comparative descriptive approach, to describe the methods of determining the different cut-off scores, in addition, to comparing the cut-off score resulting from the application of several methods.

### *Population and study sample*

The population of the study consists of the students of the educational sciences faculty at Isra University in Jordan during the academic year of 2022/2023. The sample of the study consisted of 174 students who were purposively selected to attend the test.

### *Tool*

A criterion-referenced achievement test in measurement and evaluation has been used. The test has been built according to a different phase: The content of the basics of measurement and evaluation for the university student in the Faculty of Educational Sciences was analyzed e.g., defining the measurement, evaluation, test, types of traditional measurement, types of alternative measurement, the steps of building tests, specification table, types of tests, analysis of tests, and the test psychometric properties of the test. Finally, several objectives were formulated with different levels of knowledge for every measurement subject.

### *Building of the specification table*

The specification table has been built in order of determining the relative importance of every subject, and the different levels of cognitive objectives, furthermore 45 multiple choice items have been determined.

### *Formulation of the test items*

The items of test were (45) items with four alternative multiple-choice answers, and the bases and systems of formulating the multiple-choice test have been considered. In order to examine the items' suitability, language, and the related of the items to the objective of the test, the test was checked by a group consisting of (8) professors and professionals in measurement and evaluation. The examiners' comments and feedback were considered, the items number (3, 22, 24, 29, 33, 39, and 40) have been reformulated, furthermore items number (53, 23, 7, and 44) have been deleted.

### *Analysis of the items of the test*

First, the results related to discrimination and difficulty coefficient: the test was applied to (174) participants, and the discrimination and difficulty coefficient were calculated.

Second, the indications of the test validity and reliability

- I. Content validity: the content validity was examined through the building specification table. The relative importance of every subject and the different levels of cognitive objectives were calculated, moreover, the items for each subject and cognitive objective were determined.
- II. Criterion validity: the criterion validity correlation coefficient was examined by calculating the Person correlation coefficient for the participants' marks in measurement and evaluation subjects and their marks in the final exam of natural sciences subject.
- III. Indications of the reliability: The Split Half method was run for the first half of the med exam and the second one. In addition, applying the equation of Kuder-Richardson -20.

### *Estimate the cut-off score*

The cut-off score was estimated by four methods:

- I. Angoff Method: a questionnaire was formulated consisting of the (40) items of the test, with a box in front of each item used by the examiner to estimate the probability that the student who has the minimum sufficiency to answer the items correctly without resorting to guessing. Later, the estimated probability values are collected for each item and the mean for all examiners in each item, and the mean of all items have been calculated, the cut-off score was 0.59.
- II. The Nedelsky method: after the examiners reviewed the test, their rating was determined, and the examiners asked to determine the number of students to whom the measurement and evaluation test has been applied, and who have the minimum acceptable competencies. Furthermore, identify the wrong alternatives that the students can enslave of their choice because it cannot be the correct answer to the item. Thus, the minimum level of the passage of the item is the reciprocal of the number of alternatives left for the items. Finally, Spearman's correlation coefficient has been calculated and its value was 0.64 which is an acceptable score and represents the cut-off score.
- III. Contrasting groups' method: the participants were divided according to their test mark into two groups: standardized group and unstandardized group. The standardized group was classified according to their obtained marks which were 36 and more than (those who obtained 90% and more). While the unstandardized group consisted of the participants who obtained 15 and less (those who obtained 33.3% and less). Finally, the scores of the two groups were combined and the cut-off score was 0.66.
- IV. Criterion groups' method: the participants were divided according to their test mark into two groups: standardized group and unstandardized group. The standardized group was classified according to their

obtained marks which were 32 and more than (those who obtained 80% and more). While the unstandardized group consisted of the participants who obtained 16 and less (those who obtained 40% and less). The distribution of the total scores of the two groups was represented by the graphic using the Statistical Packages in the Social Sciences (SPSS), and the cut-off score was 0.62 determined by the frequency distribution curve.

## 9. Results

Results of the first question: What are the indications of the difficulty and discrimination coefficients for the test items?

Table (1): difficulty and discrimination coefficients for the test items

Items	difficulty	discrimination	Items	difficulty	discrimination
1	0.8706	0.260	21	0.8588	0.275
2	0.7176	0.298	22	0.8824	0.188
3	0.7882	0.348	23	0.7882	0.273
4	0.5941	0.180	24	0.8471	0.508
5	0.8000	0.309	25	0.7176	0.181
6	0.8059	0.275	26	0.6235	0.450
7	0.8147	0.238	27	0.3824	0.198
8	0.7294	0.426	28	0.5882	0.446
9	0.8588	0.195	29	0.3588	0.184
10	0.5765	0.195	30	0.6706	0.226
11	0.8212	0.274	31	0.8000	0.210
12	0.7882	0.400	32	0.8235	0.357
13	0.8029	0.213	33	0.8176	0.186
14	0.8471	0.181	34	0.8824	0.492
15	0.8000	0.185	35	0.8353	0.286
16	0.8588	0.268	36	0.7882	0.424
17	0.7529	0.406	37	0.8194	0.212
18	0.8706	0.598	38	0.4059	0.194
19	0.7765	0.549	39	0.8471	0.226
20	0.8941	0.288	40	0.8000	0.204

Table (1) presented that no items have been excluded, and the values of discrimination coefficients were more than 0.18. Furthermore, the difficulty coefficient values ranged between 0.3588 to 0.8212 which indicated that the test consisting different levels of items easy, medium, and hard levels.

Results of the second question: What are the indications of the test's concurrent validity and test reliability?

First, Criterion validity: the criterion validity correlation coefficient was examined by calculating the Person correlation coefficient for the participants' marks in measurement and evaluation subjects and their marks in the final exam of natural sciences subject. The value of the criterion validity correlation coefficient was 0.76 which is indicated that the test of measurement and evaluation has a good degree of criterion validity correlation coefficient.

Second, the reliability of the test: The Spilt Half and equation of Kuder-Richardson -20 was run as shown in the below table (2):

Table (2): the reliability of the test

Method	Items	Value of reliability
Kuder-Richardson -20	40	0.81
Spilt Half	40	0.79

Results of the third question: What is the cut-off score for passing the measurement and evaluation test according to Angoff, Nedelsky, contrasting groups, and criterion groups' method?

Table (3): cut-off score according to four methods

Method	Items	Cut-off score	100%
Angoff	40	26	65%
Nedelsky	40	25	64%
Contrasting groups	40	27	68%
Criterion groups'	40	24	62%

Table (3) showed that the obtained cut-off score in four methods was convergent and valued at more than 0.50, with the values of the cut-off score obtained in four methods ranging between 0.62 to 0.68. This makes sense for a very important subject for students of Educational Sciences, as they must master several skills in the field of education.

Results of the fourth question: What is the sensitive indication of the item at the cut-off score according to

the Angoff method? To answer the question Phi coefficient ( $\phi$ ) to examine the sensitive indication of the item at the cut-off score at Angoff method as illustrated in the below table (4):

Table (4): sensitive coefficient for measurement and evaluation test at Angoff method

Items	$\phi$	Items	$\phi$	Items	$\phi$	Items	$\phi$
1	0.32	11	0.66	21	0.44	31	0.60
2	0.33	12	0.45	22	0.34	32	0.34
3	0.42	13	0.30	23	0.31	33	0.45
4	0.39	14	0.33	24	0.37	34	0.32
5	0.51	15	0.49	25	0.55	35	0.37
6	0.41	16	0.44	26	0.41	36	0.39
7	0.39	17	0.34	27	0.49	37	0.35
8	0.55	18	0.33	28	0.38	38	0.65
9	0.41	19	0.48	29	0.33	39	0.33
10	0.39	20	0.41	30	0.39	40	0.38

Table (4) indicated that all items have an acceptable level of sensitivity and the value of ( $\phi$ ) Phi coefficient at the cut-off score was more than (0.20).

#### 10. Discussion and conclusion

Based on the results obtained by the analysis of discrimination and difficulty coefficient for the test items, the test has some easy and difficult levels. While most of the tests were at a medium difficulty level. according to the results, the test was at a medium difficulty level to evaluate the individual differences between the participants, furthermore, the items of the test have been distinguished between the student with high abilities and the student with low abilities because of the value of discrimination and coefficient was more than 0.18. These results agree with the results obtained by Olimat (2020) which "indicated that item parameters estimations comprising difficulty, discrimination, and guessing were acceptable within the indexes available in educational measurement literature".

According to calculating the cut-off score of the test in several methods, the results of the study showed that the cut-off score obtained by contrasting groups' methods was the highest 0.68. The second one was the cut-off score, which is obtained by the Angoff method of 0.65, which indicated the percentage of the participants who mastered the measurement and evaluation skills was the highest. Therefore, it is preferable to use one of these two methods to determine the cut-off score in the criterion-reference test in a subject such as a measurement and evaluation, because it is a very important subject for the student in Educational Sciences, to master a basic skill such as test design, traditional and non-traditional measurement, and evaluation tools, building specifications table, test analysis, validity and reliability of the test, and descriptive statistics indicators.

Related to the concurrent Criterion validity of the test, the Pearson correlation coefficient between the marks of the final test of measurement and evaluation subject and natural sciences subject, and the correlation value was (0.76). Moreover, the reliability of the test was calculated by Cronbach Alpha and the results showed that the test enjoys a high level of reliability 0.81 and is appropriate to the target of the study, this result agrees with the results obtained by (Zamili, 2021; Al-Atwi and Al-Masoudi, 2019). For the items, the sensitivity coefficient at the cut-off score, the items have acceptable because the sensitivity coefficient of all items was more than (0.20). which means that the test items have good sensitivity to measure the skill or concept required for each item.

The current study has steps to offer novel contributions by responding to the scarcity of studies that, build a criterion-referenced test in measurement and evaluation for university students and determine its cut-off score in several methods, and its combined effect on the evaluation of the student's skills and proficiency in measurement and evaluation subject. However, several limitations need to be acknowledged for future studies to take into consideration. This study was conducted in the faculty of educational sciences at Isra University during the second semester of the 2021/2022 academic year. The results are limited to students of the Faculty of Educational Sciences and can be generalized to students of faculties of educational sciences in Jordanian universities. It is also recommended to design statistical programs to determine the cut-off score using the computer, in addition to conducting a study to predict the four best ways to determine the cut-off score in the subject of measurement and evaluation of the university student's achievement.

#### References

- Ababneh, I. (2009). Criterion-referenced test: its philosophy and the foundations of its development, Dar Al-Maserah for Publishing, Distribution and Printing, Amman, Jordan.
- Abdel-Salam, N.(1992). Multiple data for determining the interval scores in the measurement Reference Narrative (Experimental Study) Journal of the Faculty of Girls, Ain Shams University, Volume 1, Issue 1.
- Al-Atwi, H and Al-Masoudi, A . (2019). Building a criterion-reference test to measure the competencies of statistics among postgraduate students at the University of Tabuk, International Journal of Educational and

- Psychological Studies, 7(1), 97-123.
- Allam, S. (1986). Contemporary developments in psychological and educational measurement, Kuwait: Al-Qabas Commercial Press.
- Allam, S. (2000). Educational and psychological measurement and evaluation, its basics, applications and contemporary trends, Cairo, Dar Al-Fikr Al-Arabi.
- Allam, S. (2001). Educational and psychological measurement and evaluation, Cairo: Dar Al-Fikr Al-Arabi.
- Angoff, W.H. (1971), Scales, Norms and Equivalent Scores, In R.L. Thorndike (Ed.) Educational Measurement. Washington, D.C: American council on education.
- Berk, R. A . (1982). Criterion-referenced measurement. The State of Art.
- Berk, R. A. (1984). Selecting the index of reliability. A guide to criterion-referenced test construction, 231-266.
- Berk, R.A(1980). Consumers' guide to setting performance standards on the criterion-referenced test. Review of Educational Research .56:137-172.
- Ebel, R.L. (1979). Essentials of educational measurement. Englewood cliffs. Nj: prentice hall.
- Fidelis, I and Tei-Firstman, R. J. (2017). Comparison of cut-off scores using two setting methods, International Journal of Quantitative and Qualitative Research Methods, 5(2), 42 – 60.
- Gray, W. M. (1978). A comparison of Piagetian theory and criterion-referenced measurement.review of educational research ,48:223-249.
- Hambleton, R.K. (1978), On the use of cut-off score with a criterion-referenced test in an instructional setting, Jem.15:277-290.
- Meskanska, J. A. (1976). Evaluation models for criterion-referenced testing: views regarding mastery & standard setting. Review of educational research.46:133-158.
- Nunnally, J.C. (1978). Psychological theory.2nd ed.New York:McGraw-Hill.
- Olimat, M . (2022). Build a criterion-reference test in mathematics according to the three-parameter logistic model, Journal of Al-Quds Open University for Educational and Psychological Research and Studies 13(38), pp. 46-62.
- Sigmon and Haplin, G. (1983). Mini competency standards set by three divergent groups of raters using three judgmental procedures implications for validity.educational &psychological measurement, 196,43 – 185.
- Zamili, A. (2021). Using the single-parametric logistic model in building a reference test in descriptive statistics according to paragraph response theory, Journal of Basic Sciences 1(1), pp. 157-180.

#### Appendix: the test in Arabic language

ضع دائرة حول رمز الإجابة الصحيحة:		(40) علامة	
1. من أمثلة القياس غير المباشر:	(أ) الطول	(ب) الوزن	(ج) التحصيل
2. متغير الجنس (ذكر، انثى) يقع في المستوى:	(أ) الأسمي	(ب) الرتبي	(ج) الفئوي
3. واحدة مما يلي لا تعد من مزايا اسئلة الاختبار من متعدد:	(أ) المرونة	(ب) الموضوعية	(ج) الصدق
4. متغير العمر يقع في المستوى:	(أ) الأسمي	(ب) الرتبي	(ج) الفئوي
5. واحدة مما يلي لا تعد من مزايا اسئلة الاختبار من متعدد:	(أ) الشمول	(ب) الموضوعية	(ج) الصدق
6. ينصح باستخدام الأسئلة الشفوية في حالة من الحالات التالية:	(أ) إلقاء القصائد الشعرية	(ب) مع الصغار بالسن	(ج) إذا كان العدد كبير.
7. الهدف الأساسي للتقويم التشخيصي هو:	(أ) رصد العلامات	(ب) تعديل الخطط الدراسية	(ج) تحديد جوانب الضعف
8. ما اسم المصطلح الذي يعني باتخاذ القرارات ويعمل على وضع خطط العلاج:	(أ) التقويم	(ب) القياس	(ج) الاختبار
9. التقويم ذاتي المرجع يعتمد على:	(أ) مقارنة أداءه على عدد من المواد	(ب) مقارنة أدائه بطلبه من صف آخر	(ج) مقارنة أداء الطالب بنسبة تحقيقه للأهداف
10. الأسئلة المقالية الاختبارية تعني:	(أ) هي أسئلة مقالية وصح وخطأ	(ب) هي أسئلة مقالية وأكمل الفراغ	(ج) هي أسئلة مقالية وموضوعية
11. إذا كان معامل الصعوبة يساوي 0.22 فإن معامل السهولة يساوي:	(أ) 0.88	(ب) 0.78	(ج) 1
12. من فوائد جدول المواصفات المساعدة في بناء اختبار:	(أ) متوازن	(ب) مقنن	(ج) عدد فقراته قليلة
13. جميع الاختبارات التالية تصنف ضمن الاختبارات القصيرة عدا واحدة هي:	(أ) متوازن	(ب) مقنن	(ج) عدد فقراته قليلة

14. متغير عدد نبضات القلب يقع في المستوى: (ب) إملاء الفراغات (ج) التكميل (د) المقال المقيد
15. واحدة مما يلي لا تعد من مزايا الاسئلة المقالية: (ب) الرتبي (ج) الفئوي (د) النسبي
16. متغير المستوى التعليمي يقع في المستوى: (ب) الموضوعية (ج) تعطي حرية للطالب في الإجابة (د) تطبق على أعداد كبيرة
17. واحدة مما يلي لا تعتبر من عيوب الأسئلة المقالية هي: (ب) الرتبي (ج) الفئوي (د) النسبي
18. تحتاج لوقت طويل للتصحيح (ب) تخضع لذاتية المصحح (ج) لا تسمح للغش (د) تتأثر درجة الطالب بالخط والأسلوب
19. تتمتع بالموضوعية (ب) معظم الفقرات تتعلق بحقائق سطحية (ج) فيها نسبة للتخمين (د) فيها مجال للغش
20. ما أبسط مستوى من مستويات القياس الأربعة: (ب) التقييم الحقيقي (ج) التقييم التقليدي (د) التقييم البديل
21. يسمى جدول المواصفات جدولاً: (ب) الأسمي (ج) النسبي (د) الفئوي
22. واحدة مما يلي لا تعتبر من عيوب الأسئلة الشفوية هي: (ب) ثلاثي البعد (ج) ثنائي البعد (د) رباعي البعد
23. لا تسمح للغش (ب) لا توفر العدالة في توزيع الأسئلة (ج) تحتاج وقت طويل للأعداد (د) ذاتية المصحح
24. يساعد التقييم النوعي المعلم أو المعلمة التعرف على: (ب) اتجاهات الطالب (ج) تفاعله الاجتماعي مع زملائه (د) جميع ما ذكر صحيح
25. من خصائص الاختبار الموضوعية وتعني: (ب) ألا تختلف درجة الاختبار باختلاف المصحح (ج) وضوح أسئلة الاختبار (د) تغطية كامل المادة
26. جدول المواصفات يربط بين: (ب) المادة الدراسية والزمن (ج) الأهداف السلوكية واستراتيجيات التدريس (د) المادة الدراسية والأهداف السلوكية
27. تستخدم أسئلة الاختيار من متعدد في أحد المجالات: (ب) القدرة على التحليل (ج) القدرة على التطبيق (د) جميع ما ذكر
28. التقييم النوعي يعتمد في جمع معلوماته على: (ب) المقياس (ج) الاستبانة (د) الملاحظة
29. إذا كان حجم عينة طبق عليه اختبار عددها (400) فإن عدد المجموعتين المحكيتين ( العليا والدنيا ) تساوي: (ب) 27، 27 (ج) 87، 87 (د) 200، 200
30. الفقرة التي معامل صعوبتها يساوي 0.15 هي فقرة: (ب) سهلة جداً (ج) متوسطة (د) صعبة
31. أكبر قيمة لمعامل الصعوبة تساوي: (ب) 1 (ج) 0 (د) 0.5
32. الصدق التلازمي يكون بحساب معامل الارتباط بين اختبار: (ب) انجليزي وعلوم (ج) جغرافيا وتاريخ (د) تاريخ وحاسوب
33. مجموعة قيم أهداها قيمة متطرفة أفضل المقاييس استخدام هو: (ب) الانحراف المعياري (ج) المدى (د) الوسيط
34. أقل قيمة لمعامل الثبات تساوي: (ب) 1 (ج) 0 (د) 1-
35. يكون المومو أو المشتت في أسئلة الاختيار من متعدد فعال إذا كانت قيمته: (ب) موجبة (ج) 2 (د) سالبة
36. فقرة معامل صعوبتها يساوي صفر هذا يعني انها: (ب) سهلة جداً (ج) متوسطة الصعوبة (د) صعبة
37. فقرة تمييزها يساوي 0.42 هذا يعني أن الفقرة: (ب) مقبولة وبحاجة لتحسين (ج) جيدة (د) ممتازة
38. إذا كانت لديك القيم: 3، 9، 5، 8، 10 أجب عن الاسئلة من 37 إلى 40.
39. المتوسط الحسابي يساوي: (ب) 8 (ج) 9 (د) 5
40. الوسيط يساوي: (ب) 7.5 (ج) 8 (د) 8.5
41. المنوال يساوي: (ب) 7 (ج) 8 (د) لا يوجد
42. المدى يساوي: (ب) 10 (ج) 3 (د) 9