



**TEST QUALITY AND STUDENTS' PERFORMANCES:
AN APPRAISAL OF THE NATIONAL ACHIEVEMENT
TEST IN ENGLISH AND MATHEMATICS FOR
CAMEROON PRIMARY SCHOOLSⁱ**

Ernestine Wirngo Taniⁱⁱ

University of Yaounde,
Cameroon

Abstract:

Evaluation is an essential facet of education and plays a significant role in giving feedbacks to stakeholders. Pedagogy is not complete without learners' assessment and. Objective tests are used extensively as test format in primary school. However, conception and making remains a challenge to most teachers. This cast doubts over the quality. To mitigate the issues about its quality, each test format should undergo item analysis or task analysis. This study sets out to evaluate item and test quality of a national achievement test of English language using difficulty index (DIF), and discrimination indices (DI); to identify which task were appropriate for the respective levels. The study made use of data collected by the ministry of basic education aimed at measuring the true score of their learners in order to plan new pedagogic tools for improving the quality of reading and mathematics amongst primary school pupils. The Classical Test Theory (CTT) that utilizes two main statistics: the item difficulty index and the discrimination index were employed. Through an ex-post factor analysis results obtained showed that the national achievement test was easy, thus depicting the good performance of pupils' whereas in reality it is the reverse. About 90% of the pupils got Items that were virtually correct consequently useless for discriminating among pupils. Task like Measurement and size for class three, addition and subtraction and familiar word identification for class five should be completely discarded as their DIF stood at 1.00. Given that, quality control is important for test development. Teachers are recommended to perform item analysis and to synchronize classroom instruction with test items to achieve instructional validity.

Keywords: test analysis/ task analyses, difficulty/ facility index, discriminatory indices, test statistics

ⁱ QUALITE DES TESTS ET PERFORMANCES DES ELEVES : UNE EVALUATION DES TESTS DES PERFORMANCES NATIONALES EN ANGLAIS ET EN MATHEMATIQUES DES ECOLES PRIMAIRES DU CAMEROUN

ⁱⁱ Correspondence: email wirngotani@yahoo.fr

1. Background and Justification

In May 2019, the Ministry of Basic Education carried out a survey through a data collection mission to determine the real level of performance of pupils in some foundational skills: reading and mathematics aimed at improving on their general performance. These studies revealed that pupils' achievement in Mathematics and English language were good. Yet the knowledge, skills and attitudes acquired in Reading and Mathematics are very instrumental in enhancing the acquisition of knowledge, skills and attitudes in other subjects of the curriculum. It is on the bases of this that the current study seeks to find out if the exams measured what it was supposed to measure and the quality of the items and their resulting performances.

From the face value of the exams, it was evidently clear that the objective test format (True/ False, selecting, matching, fill- in, multiple choice) was used which are the most widely used to assess learners' competences (Wirngo Tani, 2019). Even though they are easy to score/ mark and grade particularly on large number of pupils, they are often technically difficult to conceive and time consuming (Mannion et al., 2018; Odukoya, Adekeye, Igbinoba, & Afolabi, 2017). Teachers are expected to be competent in constructing them and also be able to analyze whether items are valid and reliable in measuring pupil's learning achievements. According to D'Sa & Visbal Dionaldo (2017) concerns regarding the quality of tests used for assessment is now increasing. Unfortunately, only a few teachers have been trained specifically to develop quality objective test items and the skill in performing item analysis (Mannion et al., 2018). Odukoya et al. (2017), Rehman, Aslam, Hassan (2018) opines that poorly constructed items without item analysis could endanger the integrity of the entire test. Further, inaccurate evaluation could impact the grade of the pupils and create irreversible impeding consequences on the career pathway of learner (Reichert, 2011). In the study carried out by Nedeau-Cato, Laughlin, and Rus (2013), it was found that 85% of items have at least one flaw. In another related study done by Hijji (2017) examinations that did not undergo items analysis resulted to 91.8% of the items having one or more items that are flawed. It could therefore mean that conducting item analysis is essential for every exam to mitigate errors and improve integrity of each exam Item analysis is done by analyzing four components namely: Difficulty Index (DIF), Discrimination Index (DI), Distractor Efficiency (DE), and reliability test using the Kuder-Richardson Formula 20 (KR20). But this study evaluated item and test quality using difficulty index (DIF) and discrimination indices (DI).

The specific aim of this study is to find out the impact of the task analysis through its respective items on the quality of the tests and students' performances.

2. Literature Review and Theoretical Framework

Item analysis is used to ascertain the quality of an examination or a test. It enables a teacher to improve his/her competences in creating good test items. Item analysis can be

used to evaluate if the item is difficult or easy (Tracy, 2012). According to Polit and Yang (2015), item analysis is done to evaluate which items to discard, to retain, and or needs revision. Thus, an exams or test needs item analysis to measure its performance.

In an objective evaluation of pupils' performances, the item analysis is a process in which both learners' answers and test questions are examined in order to assess the quality and quantity of the items and the test as a whole. The crucial object of preparing objective tests is to construct good questions and this requires an understanding of the objectives, having good skills in writing the items and an excellent mastery of content both as a substance and as a process. Guidelines supported by experimental or quasi-experimental designs are available, but unfortunately are usually not adhered to, consequently the preparation and administration of faulty tests (Walsh, 2008; Haladyna, 2004). Item analysis is an examination of a test after its administration. The quality of a test depends upon each item of a test (Shrama, 2000). Item analysis allows us to observe the item characteristics, and to improve the quality of the test Gronlund (1993) cited in Wirngo (2018). Lange (1967) opines that, item revision allows the identification of items too difficult or too easy, items not able to differentiate between students who have mastered the content and those who have not, or questions that do not have plausible distracters. This gives teachers the opportunity to remove them from the test/ task or change the items or even modify instruction that are ambiguous, confusing and can possibly lead to misunderstanding about the content or adjust the way they are teaching. Test analysis improves skills and save time and energy of teachers and test engineers.

The theoretical approach adopted for this study is the Classical Test Theory (CTT) that utilizes two main statistics: the item difficulty index (A measure of whether an item was too easy or too hard). The range of DIF is from 0-1 and when multiplied by 100 the p-value is converted to a percentage. When the calculated value is high, it denotes that the question / item / task is easy. According to Mahjabeen et al. (2018), if the p value is between 20-90%, the question is regarded as good and acceptable and items with p-value between 40-60% are seen as excellent, items with a p - value of less than 20% is considered too difficult and more than 90% is too easy, extreme outliers which are not acceptable and need to be revise or discarded. While to Mukherjee & Lahiri (2015), items with DIF of >70% is too easy, between 30-70% is average, and between 50-60% is good. The Discrimination index (DI) (A measure of whether an item discriminated between students who knew the content well and students who did not.). The DI is used to estimate the effectiveness of an item in discriminating high achieving students from low achieving ones. In calculating the DI, the test takers are divided into quartiles. The upper quartile or students who have the highest scores, lower quartile, those who have the lowest scores, and the students who has average scores or the middle quartile. The value ranges from -1 to +1. The item is considered to be effective and is discriminating if it has a higher value (Musa, Shaheen, Elmardi, & Ahmed, 2018). Mukherjee and Lahiri (2015) explained that if all the test takers in the upper quartile and not in the lower quartile answer the item correctly, the DI value is 1.00. On the other hand, if the lower group answer it correctly and none from the upper group, the DI value would be -1.00 (D'Sa &

Visbal-Dionaldo, 2017; Mukherjee & Lahiri, 2015) maybe due to item flaws or inefficient distractors (D'Sa & Visbal-Dionaldo, 2017). According to Musa et al., (2018) these “*items should be carefully reviewed for the presence of common causes of poor discrimination such as ambiguous wording, inappropriate instructions, wrong keys and areas of controversy*”. Musa et al. (2018) concord that if the DI value is 1.00, it indicates a perfect discrimination between high and low performing students and if the value is near or less than zero, the item should be removed from the exam. This implies that, items with DI value greater than 0.40 are considered as excellent, 0.30-0.39 as reasonably good but probably needs improvement, 0.20 to 0.29 are marginal items and should be reviewed, while items with a value below 0.19 are considered poor and must be removed (Mukherjee & Lahiri, 2015). Mahjabeen et al. (2018) categorized DI as items with a value of ≥ 0.36 , excellent, between 0.25 to 0.35 as good, between 0.21 to 0.24 as acceptable, and items that are ≤ 0.20 are poor.

3. Methodology

The ex-post factor design was applied in the study that allows the researchers to analyze already existing data (Wirngo, 2019) and examine whether the instrument used was consistent and measured what it purports to measure (Houser, 2018; Polit & Yang, 2015). An after- the- fact study of marked scripts in Mathematics and Reading that was administered in twenty five government primary schools of each of the ten regions. The test items of the English language were on letter sound knowledge, familiar word reading, phonemic awareness, vocabulary, fluency and comprehension. In Mathematics the items focused on number identification, ordering, geometry, addition, subtraction, matching, ranking and tallying.

The study analyzes a national achievement exam on English language with a focused on the objective format. Each item in a task was analyzed and then the general task difficulty and discriminatory indices were computed. This population consisted of primary school pupils over the national territory for the three levels- level 1 (classes 1 & 2), level 2 (classes 3 & 4) and level 3 (classes 5 & 6).

Each item was encoded into SPSS version 20 and was analyzed using item statistics: Difficulty Index (DIF), and Discrimination Index (DI).

Table 1: Data Analyses and Interpretation

Class	Subject	N° of pupils	N° passes	N° failed	% passed
1	English	180	177	03	98.3%
	Mathematics	120	120	00	100%
2	English	180	124	56	68.9%
	Mathematics	120	82	38	68.3%
3	English	179	150	29	83.8%
	Mathematics	179	129	50	72.1%
4	English	180	170	10	94.4%
	Mathematics	180	150	30	83.3%
5	English	180	175	05	97.2%
	Mathematics	180	160	20	88.9%
6	English	180	154	26	85.6%
	Mathematics	180	176	04	97.8%

This table indicates the performance of pupils nationwide that took the test and constituted the base for item analysis.

3.1 Test Analysis per Task (Task performance)

Table 2: Class one --- Mathematics

Marks Tasks	Task-1 N° recognition 10 marks	Task-2 Ordering 7 marks	Task-3 Addition 3 marks	Task-4 Subtraction 3 marks	Task-5 Matching 5 marks	Task-6 Ranking 2 marks	Total
0	0	3	35	28	2	18	86
1	1	13	41	28	2	9	94
2	3	10	52	56	1	153	275
3	5	20	52	57	4	0	138
4	13	14	0	08	4	0	38
5	11	22	0	0	167	0	200
6	17	33	0	03	0	0	53
7	12	65	0	0	0	0	77
8	40	0	0	0	0	0	40
9	30	0	0	0	0	0	30
10	48	0	0	0	0	0	48
Total	180	180	180	180	180	180	
Observation	0.88 = 88% Easy	0.74 = 74% Right Difficulty	0.57 = 57% Right Difficulty	0.67 = 67% Right Difficulty	0.97 = 97% Easy	0.85 = 85% Easy	

The calculated item discrimination index was negative (-1) for some randomly selected items. This could be as a result of ineffective distracters as many of the lower group turn to get the items correct. Number recognition, matching and ranking tasks are found to be too easy and hence need to be revised or discarded.

Table 3: English

Marks Tasks	Letter name identification 5 mks	Letter sound identification 5 mks	Familiar words identification 5 mks	Grammar 5 mks	Total
0	3	0	0	1	4
1	14	0	0	8	22
2	29	5	14	19	67
3	34	29	50	39	152
4	24	48	65	34	171
5	76	98	51	79	304
Total	180	180	180	180	
Observation	0.74 = 74% Right difficulty	0.97 = 97% Easy	0.92 = 92% Easy	0.84 = 84% Easy	

The calculated item discrimination index was negative (-1) for some randomly selected items. This could be as a result of ineffective distracters as many of the lower group turn to get the items correct. Letter sound identification, familiar word identification and grammar tasks are found to be too easy and hence need to be revised or discarded.

Table 4: Class Two --- Mathematics

Marks Tasks	Task-1 N° recognition 10 marks	Task-2 Ordering 7 marks	Task-3 Seriation 3 marks	Task-4 Addition 3 marks	Task-5 Substruction 5 marks	Task-6 Tallying 2 marks
0	15	20	13	05	23	82
1	01	10	104	09	11	49
2	04	17	03	24	30	49
3	05	06	03	24	30	
4	11	50	15	74	64	
5	16	77	42			
6	07					
7	09					
8	10					
9	15					
10	87					
Total	180	180	180	180	180	
Observation	0.80 = 80% Easy	0.74 = 74% Right difficulty	0.33 = 33% Right difficulty	0.92 = 92% Easy	0.87 = 87% Easy	0.54 = 54% Right difficulty

The calculations in the table above indicate that the item discrimination index was negative (-1) for some randomly selected items. This could be as a result of ineffective distracters as many of the lower group turn to get the items correct. Number recognition, addition and subtraction tasks are found to be too easy and hence need to be revised or discarded.

Table 5: English

Marks Tasks	Letter name identification 5 mks	Letter sound identification 5 mks	Familiar words identification 5 mks	Grammar 5 mks	Listening Comp. 5 mks
0	00	12	28	05	31
1	16	05	06	11	12
2	22	08	13	19	14
3	06	10	14	27	36
4	09	17	17	78	47
5	67	128	102	40	40
Total	120	180	180	180	180
Observation	0.68 = 68% Right difficulty	0.86 = 86% Easy	0.74 = 74% Right difficulty	0.81 = 81% Easy	0.68 = 68% Right difficulty

Looking at the table above, it can be observed that the calculated item discrimination index was negative (-1) for some randomly selected items. This could be as a result of ineffective distracters as many of the lower group turn to get the items correct. Letter sound identification and grammar tasks are found to be too easy and hence need to be revised or discarded.

Table 6: Class Three --- Maths

Marks Tasks	Task-1 Addition 4 marks	Task-2 Subtraction 4 marks	Task-3 Geometry 4 marks	Task-4 M/S (time) 4 marks	Task-5 M/S (length) 4 marks	Task-6 Gr/Stat 4 marks
0	4	21	24	109	165	15
1	18	38	23	70	14	24
2	55	38	19	/	/	57
3	38	40	9	/	/	77
4	64	42	104	/	/	6
Total	179	179	179	179	179	179
Observations	0.87 = Easy	0.66 = Right difficulty	0.66 = Right difficulty	1.00 Easy	1.00 = Easy	0.78 = Easy

Based on the calculations above, the calculated item discrimination index was negative (-1) for some randomly selected items. This could be as a result of ineffective distracters as many of the lower group turn to get the items correct. Addition, measurement and size (time), measurement and size (length) and graphs and statistic tasks are found to be too easy and hence need to be revised or discarded.

Table 7: English

Marks Tasks	Letter name identification 5 mks	Letter sound identification 5 mks	Familiar words identification 5 mks	Reading Comprehension 5 mks	Active Listening/ 5 mks	Grammar/ 5 mks
0	47	11	0	15	21	1
1	80	19	14	16	11	7
2	27	20	12	16	34	13
3	11	45	30	42	33	20
4	06	42	40	59	51	50
5	08	42	83	31	29	88
Total	179	179	179	179	179	179
Observation	0.13 = Difficult	0.72 = Right difficulty	0.85 = Easy	0.73 = Right difficulty	0.63 = Right difficulty	0.88 = Easy

Analysis from the tests above indicates the calculated item discrimination index was negative (-1) for some randomly selected items. This could be as a result of ineffective distracters as many of the lower group turn to get the items correct. Letter name identification task is too difficult while familiar word identification and grammar tasks are found to be too easy and hence all tasks need to be revised or discarded.

Table 8: Class Four --- Mathematics

Marks \ Tasks	Task-1 Ranking 5 marks	Task-2 Addition 4 marks	Task-3 Subtraction 4 marks	Task-4 M/S Time 1 mark	Task-5 Operations 3 marks	Task-6 Geometry 2 marks
0	12	02	00	157	43	175
1	04	06	19	23	47	05
2	07	11	25		40	
3	21	38	53			
4	47	123	83			
5	89	01	00			
Total	180	180	180	180	180	180
Calculated DIF	0.88	0.96	0.89	0.13	0.55	0.03
Observation	Easy	Easy	Easy	Difficult	Right difficulty	Difficult

Looking at the table above, it can be observed that the calculated item discrimination index was negative (-1) for some randomly selected items. This could be as a result of ineffective distracters as many of the lower group turn to get the items correct. Ranking, addition and subtraction tasks are found to be too easy while measurement and size (time) and geometry task are found to be too difficult and hence these tasks need to be revised or discarded.

Table 9: English

Marks \ Tasks	Letter sound identification 5 marks	Familiar words identification	Reading comprehension 5 marks	Active Listening 5 marks	Vocabulary 5 marks	Grammar 10 marks
0	00	00	08	08	17	03
1	02	09	10	09	11	04
2	03	09	29	26	13	14
3	09	16	39	36	29	15
4	51	52	54	42	68	31
5	115	94	40	34	42	113
Total	180	180	180	155	180	180
Calculated DIF	0.97	0.90	0.74	0.72	0.77	0.88
Observation	Easy	Easy	Right difficulty	Right difficulty	Easy	Easy

Looking at the table above, it can be observed from the analysis that the calculated item discrimination index was negative (-1) for some randomly selected items. This could be as a result of ineffective distracters as many of the lower group turn to get the items correct. Letter sound identification, familiar word identification, vocabulary and grammar tasks are found to be too easy and hence need to be revised or discarded.

Table 10: Class Five --- Mathematics

Tasks \ Marks	Task-1 Sets/L 5marks	Task-2 Add. /Subtr 5 marks	Task-3 Operation 3 marks	Task-4 Geometry 3 marks	Task-5 M/S length 1 mark	Task-6 M/S Time 3marks
0	16	0	6	6	40	69
1	34	0	31	33	140	72
2	53	0	48	84	0	15
3	50	14	95	57	0	24
4	01	56	0	0	0	/
5	22	110	/	/	0	/
Total	180	180	180	180	180	180
Calculated DIF	0.40	1.00	0.79	0.78	0.77	0.21
Observation	Right difficulty	Easy	Easy	Easy	Easy	Difficult

From the table above, it can be observed that the calculated item discrimination index was negative (-1) for some randomly selected items. This could be as a result of ineffective distracters as many of the lower group turn to get the items correct. Addition & subtraction, operations, geometry, measurement and size (length) tasks are found to be too easy while measurement & size (time) is found to be a very difficult task and hence, all these tasks need to be revised or discarded.

Table 11: English

Tasks \ Marks	Familiar words identification 5 marks	Reading comprehension 5 marks	Active Listening 5 marks	Grammar 5 marks	Vocabulary 5 marks	Directed Writing 10 marks
0	0	0	02	0	01	05
1	0	04	11	01	03	05
2	0	5	13	01	01	33
3	04	25	23	11	07	05
4	15	66	42	36	02	20
5	161	80	89	131	146	0
6	0	0	0	0	0	38
7	0	0	0	0	0	0
8	0	0	0	0	0	36
9	0	0	0	0	0	0
10	0	0	0	0	0	38
Total	180	180	180	180	180	180
Calculated DIF	1.00	0.95	0.85	0.98	0.86	0.76
Observation	Easy	Easy	Easy	Easy	Easy	Easy

From the above analysis, it is observed that English language was generally easy for class five pupils. With the aid of the taxono-metric levels and the curriculum, test constructors are advised to review the different items of the various tasks to include higher levels of achievement. In this way item discrimination can be improved upon.

Table 12: Class Six --- Maths

Marks Task	Task-1 Statistics/ 5mks	Task-2 Operations/ 4mks	Task-3 Geometry/ 4mks	Task-4 M/S time/1mk	Task - 5 Gr/Stat/ 3mks	Task - 6 Ordering/ 2mks
0	02	06	25	22	28	47
1	03	26	09	120	40	27
2	09	15	20	05	35	106
3	40	50	33	33	77	00
4	90	75	93	00	00	00
5	36	08	00	00	00	00
Total	180	180	180	180	180	180
Calculated DIF	0.92	0.82	0.81		0.62	0.73
Observations	Easy	Easy	Easy		Right difficulty	Right difficulty

From the analysis in the table above, it can be observed that the calculated item discrimination index was negative (-1) for some randomly selected items. This could be as a result of ineffective distracters as many of the lower group turn to get the items correct. Statistics, operations and geometry tasks are found to be too easy and hence need to be revised or discarded.

Table 13: Class Six --- English

Marks Tasks	Reading Comp./ 5mks	Vocabulary/ 5 mks	Grammar 5 mks	Active Listening/ 5mks	Composition Writing/ 20mks
0	01	0	0	01	
1	02	01	01	03	
2	04	02	02	18	
3	65	08	07	23	
4	107	29	39	50	
5		140	131	85	03
6					03
7					04
8					07
9					0
10					08
11-13					79
14-16					50
17-20					18
Total	180	180	180	180	180
Calculated DIF	0.95	0.98	0.98	0.87	0.86
Observation	Easy	Easy	Easy	Easy	Easy

From the above analysis it is observed that the test as a whole was very easy for class six. The entire test should be reviewed and some challenging exercises of the three higher levels of cognition included.

4. Discussion

From the analysis, it is evident that exams lacked the right difficulty level and had a poor high discrimination power. About 80% of the items DIF were easy with poor discrimination indices; the findings of this study contrast with some related research in which the item difficulty is considered an acceptable level with a high discrimination level (Mahjabeen et al. 2017; Mukherjee & Lahiri, 2015). Consequently, teacher should also rethink measuring the efficiency of the distractors of each item. Miller et al (as cited in Oermann and Gaberson, 2014), opine that, educators should be careful in deleting items with poor results in DIF and DI because it could negatively impact the validity of the exam due to fewer sample content. From all indications, a complete review is necessary for class three measurement and size (time & length), class five addition and subtraction and familiar word identification because their calculated *P value stood at 1.00*. Even though the average DI and DIF results are not desirable, this achievement test entirety is considered “not good” with reservation because the Distracter analysis was not done, meaning that items should be moderated carefully when deciding whether an item should be included or not in the post exams even if the DIF and DI are not “ideal”. Item analysis will not be complete without the analysis of the distractors because the presence of distractors itself enhances the measurement properties of each item. Burud, Nagandla, and Agarwal (2019), opined that, one important feature of quality assurance of an examination is through the implementation of test analysis.

5. Conclusion

This study set out to ascertain quality assurance of examination, through a task analysis of national English and mathematics achievement exams administered in 240 Government schools. From the findings it is recommended that teachers should improve upon their test construction practices and be trained on test analysis procedures in order to ascertain quality assurance of examinations for non-standardized and standardized purposes.

Conflict of Interest Statement

I declare no conflict of interest on this article. Rather the in-depth analysis of data will be of interest to my former ministry to review their assessment practices.

About the Author

Ernestine Wirngo Tani is currently a permanent lecturer of the Higher Teacher Training College (ENS) Yaounde 1. She is a Ph.D. holder in Curriculum and Evaluation. Option - Measurement and Evaluation; Specialty: Evaluation and Docimology obtained from the Faculty of Education, University of Yaounde 1. She is also holder of a Master's and Bachelor's Degrees in Measurement & Evaluation and Developmental psychology respectively from the same University. She is a member of the National Team on Action

Research for the Measurement of Beneficiaries of Functional Literacy Programmes and an expert in Measurement and Evaluation of competences of learners in the Ministry of Basic Education. She is a seasoned pedagogue who has moved from the Classroom to a pedagogic animator at the inspectorate and regional levels; a one-time divisional Adviser for Bilingualism and a National Pedagogic staff of the ministry of Basic Education-Cameroon before her recruitment in the University of Yaounde 1.

References

- Barud, I., Nagandla, K., & Agarwal, P. (2019). Impact of distractors in item analysis of multiple-choice questions. *International Journal of Research in Medical Sciences*, 7(4), 1136-1139. doi: <http://dx.doi.org/10.18203/2320-6012.ijrms20191313>
- D'Sa, J. L. & Visbal-Dionaldo, M. L. (2017). Analysis of multiple choice questions: Item difficulty, discrimination index and distractor efficiency. *International Journal of Nursing Education*, 9(3), 109-114. doi:10.5958/0974-9357.2017.00079.4.
- Gronlund, N. E. (1998). *Assessment of student achievement*. 6th edition. Boston: Allyn and Bacon.
- Haladyna, T. M. (2004). MC formats. In: Haladyna TM (ed). *Developing and validating multiple-choice test item*. 3rd edn. Mahwah, New Jersey: Lawrence Erlbaum Associates; 67±96.
- Hijji, B. M. (2017). Flaws of multiple choice questions in teacher-constructed nursing examinations: A pilot descriptive study. *Journal of Nursing Education*, 56(8), 490-496. doi: 10.3928/01484834-20170712-08
- Houser, J. (2018). *Nursing research: Reading, using, and creating* (4th ed.). Burlington, MA: Jones & Bartlett Learning.
- Lange, A., Lehmann, I. J. & Mehrens, W. A. (1967). Using item analysis to improve tests. *Journal of Educational Measurement*, 4(2), 65-68; <http://www.jstor.org/stable/1434299>.
- Longman. Walsh, K. (2008). Advice on writing multiple choice questions (MCQs). *BMJ Careers* 2005. Available at <http://careers.bmj.com/careers/advice/view-article.html?id=616>, accessed on 23 May 2008.
- Mahjabeen, W., Alam, S., Hussan, U., Zafar, T., Butt, R. Konain, S., & Rizvi, M. (2018). Difficulty index, discrimination index, and distractor efficiency in multiple choice questions. *Annals of Pakistan Institute of Medical Sciences*, 4,310-315. Retrieved from https://www.researchgate.net/publication/323705126_Difficulty_Index_Discrimination_Index_and_Distractor_Efficiency_in_Multiple_Choice_Questions
- Mannion, C. A., Hnatyshyn, T., O'Rae, A., Beck, A. J., & Patel, S. (2018). Nurse Educators and Multiple-Choice Examination Practices. Retrieved from <http://hdl.handle.net/1880/108887>

- Mukherjee, P. & Lahiri, S. K. (2015). Analysis of multiple-choice questions (MCQs): Item and Test Statistics from an assessment in a medical college of Kolkata, West Bengal. *Journal of Dental and Medical Sciences*, 14(12), 47-52. www.iosrjournals.org.
- Musa, A., Shaheen, S., Elmardi, A., & Ahmed, A. (2018). Item difficulty & item discrimination as quality indicators of physiology MCQ examinations at the Faculty of Medicine, Khartoum University. *Khartoum Medical Journal*, 11(02), 1477-1486. Retrieved from https://www.researchgate.net/publication/328583573_Item_difficulty_item_discrimination_as_quality_indicators_of_physiology_MCQ_examinations_at_the_Faculty_of_Medicine_Khartoum_University
- Nedeau-Cayo, R. (2013). Assessment of item-writing flaws in multiple-choice questions. *Journal for Nurses in Professional Development*, 29(2), 52-57. doi:10.1097/NND.0b013e318286c2f1
- Odukoya, J. A., Adekeye, O., Igbino, A. O., and Afolabi, A. (2017). Item analysis of university-wide multiple-choice objective examinations: the experience of a Nigerian private university. *European Journal of Methodology*, 52(3), 983-997. doi: <https://doi.org/10.3928/01484834-20170712-08>
- Oermann, M. H. & Gaberson K. B. (2014). *Evaluation and testing in nursing education* (4th ed). New York, NY: Springer Publishing Company.
- Polit, D. F. & Yang, F. M. (2015). *Measurement and the measurement of change*. Philadelphia: Wolters Kluwer.
- Popham, J. W. (2008). *Classroom assessment: What teachers need to know* (5th ed.) Boston, MA: Pearson Education, Inc. Research into Higher Education and Open University Press.
- Rehman, A., Aslam, A. & Hassan, S. H. (2018). Item analysis of multiple choice questions. *Pakistan Oral and Dental Journal*, 38(2), 291-293. Retrieved from <https://www.podj.com.pk/index.php/podj/article/view/245>
- Reichert, T. G. (2011). Assessing the use of high-quality multiple-choice exam questions in undergraduate nursing education: Are educators making the grade? Retrieved from Sophia, the St. Catherine University repository. https://sophia.stkate.edu/ma_nursing/15
- Remmers H. H., Gage, N. L. & Rummel, J. I. (1967). *A Practical introduction to measurement and evaluation* (2nd ed.). Delhi: Universal Book Stall
- Sharma, S. R. (2000). *Modern teaching strategies*. New Delhi: Omsons Publications.
- Swanson, D. B., Holtzman K. Z., Albee K, & Clauser, B. E. (2006). Psychometric characteristics and response times for content-parallel extended-matching and one-best-answer items in relation to number of options. *Acad Med*, 81, 52±55.
- Swanson, D. B., Holtzman K. Z., Clauser, B. E., Sawhill, A. J. (2005). Psychometric characteristics and response times for one-best-answer questions in relation to number and sources of options. *Acad Med*, 80, 93-96.

- Tracy, D. A. (2012). *School improvement: Revitalize your school with strategic planning*. USA: Xlibris Corporation.
- Walsh, K. (2008). Answering multiple choice questions. *BMJ Careers* 2005. Available at <http://careers.bmj.com/careers/advice/view-article.html?id=891> (accessed on 23 May 2008).
- Wirngo T. E. (2019). *Evaluation processes and examination irregularities among students in Cameroon Higher Education*. Ph.D. Thesis, defended in the Faculty of Education, University of Yaoundé 1- Cameroon.

Creative Commons licensing terms

Author(s) will retain the copyright of their published articles agreeing that a Creative Commons Attribution 4.0 International License (CC BY 4.0) terms will be applied to their work. Under the terms of this license, no permission is required from the author(s) or publisher for members of the community to copy, distribute, transmit or adapt the article content, providing a proper, prominent and unambiguous attribution to the authors in a manner that makes clear that the materials are being reused under permission of a Creative Commons License. Views, opinions and conclusions expressed in this research article are views, opinions and conclusions of the author(s). Open Access Publishing Group and European Journal of Education Studies shall not be responsible or answerable for any loss, damage or liability caused in relation to/arising out of conflicts of interest, copyright violations and inappropriate or inaccurate use of any kind content related or integrated into the research work. All the published works are meeting the Open Access Publishing requirements and can be freely accessed, shared, modified, distributed and used in educational, commercial and non-commercial purposes under a [Creative Commons Attribution 4.0 International License \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/).