



Development and Preliminary Validation of Difficulty Index, Discrimination and Distractor of Measurement and Evaluation Proficiency Test for Undergraduate Students

*Kaku Zakka¹, Nkechi P.M. Esomonu², Comfort Kwatambia Olamide¹

¹Department of Social Science Education, University of Maiduguri

²Department of Educational Foundations, Nnamdi Azikiwe University Awka

*Corresponding Email: kakuzakka@gmail.com

Abstract

Measurement and evaluation proficiency test for assessing University undergraduate students was developed and validated in this study. The study was guided by Five objectives: the difficulty indices of the test items in proficiency test measurement and evaluation proficiency test (PTMEAT); the discrimination indices of the PTME items; the distracter indices of the PTME items; to determine the reliable extent of PTME items; estimating the validity of PTME items, guided the study. One hundred (100) undergraduate students of University of Maiduguri were used for the pilot testing and reliability index respectively. A sample of seven hundred (700) undergraduate students were randomly drawn for the study. One hundred (100) items with difficulty indices ranged from 0.30 to 0.70 and discrimination indices of 0.20 to 0.50 were retained. Face and content validation of the instrument was ensured by two experts in measurement and evaluation. The reliability estimates were established through test re-test method and gave a coefficient of 0.82. The test was found to be of good quality, valid and highly reliable. The Proficiency test is therefore recommended for use in assessing undergraduate students in measurements and evaluation and similar test items should be developed in all the education courses for University Undergraduates.

Keywords: Development, Validation, Standardization, Measurement and Evaluation Proficiency-Test.

Introduction

Measurement and evaluation courses equip students with skills to design tests, interpret scores, and use data for instructional improvement studies. Nitko (2001), emphasis the need for valid assessment to ensure graduate can apply these competencies in real world setting.

Most university assessments rely on teacher-made test lacking psychometric rigor. Few studies (Oduor & Kibul, 2018) have developed validated measurement and evaluation proficiency scale for undergraduates, especially in African contexts. The gap underscores the need for a locally relevant, psychometrically sound instrument.

Measurements and Evaluation are of great importance in determining the extent to which objectives have been achieved or revealing the success of teaching. In addition to being an essential component in the evaluation of educational programmes, measurements and evaluation have also been utilized as a tool at the point of transaction between educational levels, admittance to employment, and educational institution (Baykal, 2011). In addition, the teacher strategy document emphasized the issue of training teachers who will carry out measurement and evaluation activities according to some standards (MNE, 2017b). Moreover, in the education vision 2023 Document, measurement and evaluation draw attention as one of the main issues that should be emphasized for the education system's success (MNE, 2019). These points indicated the importance of measurements and evaluation as a teaching qualification by the institutions responsible for teacher training and empowerments. This situation also draws attention to the quality of teaching activities expected to provide teachers with measurements and evaluation competencies (Brookhart, 2011; Popham, 2017).

In the teaching-learning environment, there is a constant need to gauge the outcome of the quality of responsiveness of the teaching and learning process (Brookhart, 2011; Popham, 2017). This important symbiotic process generally referred to as assessment does not only occur after teaching but can also be undertaken before teaching is affected or during the teaching process (Brown, 2004; Nitko & Brookhart, 2014).



The purpose of a testing is to determine whether undergraduate students have learned what they were expected to learn or the extent to which the student have learned. They may be used to measure learning progress and proficiency and to evaluate the effectiveness of educational programmes (AERA, APA, & NCME, 2014; Linn & Gronlund, 2000).

Proficiency test is an exam which test how proficient or skilled someone is in a particular activity, field of study. Proficiency testing test accuracy that the students can achieve. Proficiency testing is not limited to any one course, curriculum or single skill rather, it tests the overall ability. Proficiency test has traditionally consisted of standardized multiple-choice items (Alderson, 2005; Brown & Abeywickrama, 2019).

The purpose of any proficiency test is to find out whether the researcher have already got the knowledge and the skills that are taught in a particular course even though the researcher has not taken the course on the campus. If the student does well on the proficiency test, then they can earn academic credit in the subject, and it is assumed you are prepared to succeed in the subject's subsequent courses.

Sufficiently high scores on the proficiency tests that are offered by the departments can also earn credit. Departmental proficiency exams are usually administered on the campus either the week before classes start or during the first week of classes in both the fall and the spring semesters. NOTE: Most of the departmental language proficiency exams do not award course credit in the language, but strong performance does fulfill the General Education language requirement.

In this study, the measurement and evaluation proficiency was developed and validated test which can be used to determine the proficiency of undergraduate education students in university. Multiple-choice question (MCQs) was chosen because it is versatile and can be used to measure all levels of cognitive. Multiple-choice question (MCQs) was used increasingly due to their reliability, validity and ease of scoring (Mehta and Mokohais, 2014; Haladyna, Downing & Rodriguez, 2002).

Esomonu and Agbonkpolo (2010) and Osadebe (2012) observed that most teachers are not good in constructing valid and reliable test in their various subject areas. Teachers find it easy to construct test items in the lower cognitive levels (knowledge and comprehension) than the higher cognitive levels (application, analysis, synthesis and evaluation). This constitutes an educational problem. Therefore, there is a need for valid and reliable test items for evaluating undergraduate students in measurement and evaluation. Validation involves a couple of steps, scrupulous, analysis, and substantial time (Esomonu and Agbonkpolo, 2010; AERA, APA, & NCME, 2014). Therefore, there is a dire need for experts to construct enough valid and reliable proficiency tests for use in University for Undergraduate students.

Ofon U. J., et al. (2022) also observed development, validation and standardization of instrument to measure statistical literacy among universities undergraduates in Cross River state. The study developed, validated, and standardized a statistical literacy test for university undergraduates in Cross River State, Nigeria. The result showed that only 25 items survived the last item assessment. The instrument was found to have strong construct validity as the items in the correlation matrix were all positively correlated. The statistical literacy scale showed convergence with statistical reasoning but diverged with statistical anxiety. The reliability of the instrument was found to be high as the coefficients of the scale were 0.81, 0.77, and 0.78 respectively, and it was adjudged reliable. Based on the findings of the study, it was concluded that the instrument was valid and reliable (Ofon et al., 2022). Thus, it was recommended that students' skills in statistics should not just be on the computational aspect but on the utilization of statistical knowledge. Statistics lecturers should emphasize statistical literacy to help students reason and think statistically, and the instrument should be widely distributed for use by researchers and psychometricians (Ofon et al., 2022).

Purpose of the Study

The Study developed, validated and standardized proficiency test in measurement and evaluation for undergraduate students ascertained:

1. the difficulty indices of the test items in proficiency test measurement and evaluation proficiency test (PTMEAT).



2. the discrimination indices of the PTME items.
3. the distracter indices of the PTME items.
4. to determine the reliable extent of PTME items.
5. estimating the validity of PTME items.

Methodology

Test Development Process

The test items were generated from university of Maiduguri; ATBU Bauchi; Bayero University Kano and Nnamdi Azikiwe University Awka four academic session past examination questions. The item was subjected to validation by professors of measurement and evaluation from University of Maiduguri and Nnamdi Azikiwe Awka. Professors were chosen because of their years of experience in the field and teaching experience in measurement and evaluation, while one senior lecturer from university of Maiduguri was chosen based on year of experience too. In the process of getting the item used for the study, initially, one hundred and twenty items were selected, after validation it was reduced to one hundred items.

The stages of proficiency test development used in this study as listed in Nkechi Esomonu Instrumentation Model (Esomonu, 2021) comprises of initial steps, test blue print, item generation, preliminary validation, trial testing, item analysis, good item selection, final testing, norming and manual and dissemination/publication. The following describes each of the stages:

A total of 700 undergraduate students of University Maiduguri were randomly selected from all the students of Education University of Maiduguri Borno state, Nigeria as the piloting study. The test content area was based on the undergraduate course outlines for Test and Measurement.

Preliminary validation: A test to measure proficiency in measurement and evaluation among part 3 undergraduate students was developed and validated. The test item was made up 100 multiple-choice test items constructed based on course outlines of part 3 undergraduate (EDU 302) course. Preliminary validation was done by three experts in measurement and evaluation in Nnamdi Azikiwe University and University of Maiduguri (two professors) in measurement and evaluation. The pilot testing was conducted to ensure compatibility of the test items while test try: out 700 undergraduate part 3 Education students yielded the data for item analysis. Analysis were attained through the tool SPSS. Initially one hundred and twenty test item were constructed, face and content validation by three experts in educational measurement, evaluation and research from Nnamdi Azikiwe University and University of Maiduguri. The expert's observations, comments, critique and suggestions were used in the modifications of the test.

Trial testing: In the preliminary stage, the 120 multiple-choice test items constructed were pilot tested on undergraduate students of faculty of Education which was not included in the main study which was used for item for the item analysis.

An item was considered good for inclusion in the final draft of the test if it had difficulty index of 0.30 to 0.70, discrimination index greater than 0.20 and a positive distracter index.

The undergraduate students' responses to the one hundred and twenty proficiency test in Measurement and Evaluation Proficiency Test (PTMEAT) were subjected to item analysis using Item Response Theory (IRT) to ensure the difficulty index, discriminative index and distractor index of the items. Item Response Theory (IRT) psychometric analysis were done to estimate the test items parameters using the information and data obtained from the undergraduate students from University of Maiduguri.

Good items selection: The one hundred items retained in the final draft of the test was based on difficulty index. The difficulty indices of the one hundred and twenty were computed using the responses of the 100 undergraduates' students of University of Maiduguri. Thereafter, one hundred items whose difficulty indices fell between 0.3 and 0.7 were selected to make up the final items that were used.



Final testing: The final version of the proficiency test in Measurement and Evaluation (100) were arranged in line with the content area of each sub-topic of the course outline.

Difficulty and discrimination indices of PTMEAT items

The results of difficulty and discriminative indices of Proficiency Test in Measurement and Evaluation Proficiency Test (PTMEAT) Item Analysis.

Table 1: The difficulty and discrimination indices obtained after PTMEAT items Analysis

Items	Key (K)	No of correct responses among higher achievers (H) N=233	No of correct responses among lower achievers (L) N=233	Difficulty index $P = \frac{Nu + Nt}{2N}$	Discrimination index $\frac{Nu + Nt}{N}$	Remark
1	A	120	46	0.36	0.32	Retain
2	C	82	40	0.26	0.18	Reject
3	B	180	68	0.53	0.48	Retain
4	B	140	56	0.42	0.36	Retain
5	D	180	162	0.73	0.08	Reject
6	A	126	52	0.38	0.32	Retain
7	D	150	100	0.54	0.21	Retain
8	C	142	100	0.52	0.18	Reject
9	A	200	150	0.75	0.21	Retain
10	C	99	50	0.32	0.21	Retain
11	B	110	45	0.33	0.28	Retain
12	A	140	72	0.45	0.29	Retain
13	C	150	50	0.43	0.43	Retain
14	D	176	60	0.51	0.50	Retain
15	B	125	60	0.40	0.28	Retain
16	B	130	48	0.38	0.35	Retain
17	D	140	20	0.34	0.52	Retain
18	A	147	86	0.05	0.26	Retain
19	A	180	122	0.65	0.25	Retain
20	A	130	12	0.30	0.51	Retain
21	B	91	41	0.28	0.21	Retain
22	C	80	28	0.23	0.22	Retain
23	A	112	28	0.30	0.36	Retain
24	B	132	62	0.42	0.30	Retain
25	C	148	82	0.49	0.28	Retain
26	C	50	22	0.15	0.12	Reject
27	C	103	46	0.32	0.24	Retain



28	D	140	48	0.40	0.39	Retain
29	D	130	106	0.51	0.10	Reject
30	C	120	70	0.41	0.21	Retain
31	B	132	84	0.46	0.21	Retain
32	B	122	142	0.57	-0.09	Reject
33	A	114	54	0.36	0.26	Retain
34	B	124	30	0.33	0.40	Retain
35	C	136	66	0.43	0.30	Retain
36	A	150	100	0.54	0.21	Retain
37	D	132	62	0.42	0.30	Retain

38	C	116	44	0.34	0.31	Retain
39	A	120	40	0.34	0.34	Retain
40	D	111	62	0.37	0.21	Retain
41	A	156	86	0.52	0.30	Retain
42	C	116	60	0.38	0.24	Retain
43	A	104	46	0.32	0.25	Retain
44	D	122	50	0.37	0.31	Retain
45	C	40	20	0.13	0.09	Reject
46	B	100	70	0.36	0.24	Retain
47	D	146	90	0.51	0.24	Retain
48	D	140	106	0.53	0.15	Reject
49	A	120	44	0.35	0.33	Retain
50	B	142	100	0.52	0.61	Retain
51	C	100	26	0.27	0.32	Retain
52	A	138	46	0.39	0.39	Retain
53	B	100	45	0.31	0.24	Retain
54	D	112	44	0.35	0.25	Retain
55	C	127	70	0.42	0.24	Retain
56	A	170	72	0.52	0.42	Retain
57	B	156	82	0.51	0.32	Retain
58	C	146	80	0.48	0.28	Retain
59	D	118	68	0.40	0.21	Retain
60	A	115	60	0.38	0.24	Retain
61	A	138	30	0.36	0.46	Retain
62	C	180	100	0.60	0.34	Retain
63	A	84	72	0.33	0.05	Reject
64	B	90	26	0.24	0.27	Retain
65	A	120	46	0.36	0.32	Retain
66	B	129	60	0.41	0.30	Retain



67	B	26	46	0.15	-0.09	Reject
68	D	120	50	0.36	0.30	Retain
69	B	108	74	0.40	0.15	Reject
70	D	166	60	0.48	0.45	Retain
71	B	166	62	0.49	0.45	Retain
72	D	114	32	0.31	0.35	Retain
73	C	140	88	0.49	0.22	Retain
74	D	126	42	0.36	0.35	Retain
75	C	130	52	0.39	0.33	Retain
76	B	136	50	0.49	0.37	Retain
77	C	134	38	0.37	0.41	Retain
78	D	130	60	0.41	0.30	Retain
79	D	120	36	0.33	0.36	Retain
80	D	164	100	0.55	0.27	Retain
81	A	146	68	0.46	0.33	Retain
82	D	122	61	0.39	0.26	Retain
83	D	99	42	0.30	0.24	Retain
84	C	108	74	0.39	0.15	Reject
85	B	152	98	0.54	0.23	Retain
86	B	08	16	0.05	-0.03	Reject
87	C	140	42	0.39	0.42	Retain
88	C	58	46	0.22	0.05	Reject
89	B	174	50	0.48	0.53	Retain
90	A	158	104	0.56	0.23	Retain
91	A	138	78	0.46	0.26	Retain
92	A	117	32	0.32	0.36	Retain
93	C	200	32	0.50	0.72	Retain
94	B	142	90	0.50	0.22	Retain
95	B	132	96	0.49	0.15	Reject
96	C	94	81	0.38	0.06	Reject
97	D	158	46	0.44	0.48	Retain
98	C	152	86	0.51	0.28	Retain
99	C	101	32	0.28	0.30	Retain
100	B	124	36	0.34	0.38	Retain
101	B	102	32	0.29	0.30	Retain
102	D	96	28	0.27	0.29	Retain
103	A	102	70	0.37	0.14	Reject
104	C	140	70	0.45	0.30	Retain



105	B	66	28	0.20	0.16	Reject
106	C	140	28	0.36	0.48	Retain
107	B	107	36	0.30	0.30	Retain
108	B	123	72	0.42	0.23	Retain
109	A	66	12	0.17	0.23	Reject
110	B	148	81	0.49	0.29	Retain
111	D	122	106	0.49	0.07	Reject
112	A	136	66	0.43	0.30	Retain
113	D	157	27	0.39	0.56	Retain
114	B	126	34	0.34	0.39	Retain
115	D	96	54	0.32	0.18	Reject
116	C	122	32	0.32	0.39	Retain
117	A	127	60	0.40	0.29	Retain
118	D	100	42	0.30	0.25	Retain
119	B	87	60	0.32	0.12	Reject
120	C	181	40	0.47	0.61	Retain



Distractor indices of the PTMEAT

The results of Distractor indices of Proficiency Test in Measurement and Evaluation Proficiency Test (PTMEAT) Item Analysis

Table 2: Distractor Indices of PTMEAT



Item	No of respondents who choose option A,		No of respondents who choose option B,		No of Respondents who choose option C		No of respondents who choose option D		Omit	Total	Distractibility $\frac{L-H}{U}$		Index $\bar{N} = 233$		Remarks
NO	L	H	L	H	L	H	L	H			A	B	C	D	
1	30	158	60	22	76	21	57	20	22	466	K	0.16	0.24	0.16	A good item
2	64	10	28	06	58	161	81	40	18	466	0.23	0.09	K	0.18	A good item
3	70	20	43	176	50	10	60	23	14	466	0.21	K	0.17	0.16	A good item
4	63	42	61	38	58	26	41	107	30	466	0.09	0.1	0.14	K	A good item
5	92	47	60	06	26	10	28	160	37	466	0.19	0.23	0.07	K	A good item
6	36	136	60	20	46	30	71	43	24	466	K	0.17	0.07	0.12	A good item
7	50	10	70	36	80	53	23	124	20	466	0.17	0.15	0.12	K	A good item
8	55	17	48	48	40	118	60	30	50	466	0.16		K	0.13	
9	20	132	06	08	60	56	69	27	88	466	K	-0.01	0.01	0.18	Option B replaced
10	60	32	61	28	48	108	74	45	10	466	0.12	0.14	K	0.12	A good item
11	52	12	34	128	63	36	64	35	42	466	0.17	K	0.16	0.12	Distractors are ok
12	36	140	64	30	60	30	63	23	20	466	K	0.15	0.13	0.17	Distractors are ok
13	44	23	58	22	50	141	61	37	30	466	0.09	0.15	K	0.1	A good item
14	70	25	50	43	40	119	63	36	20	466	0.19	0.03	0.12	K	A good item
15	50	06	50	116	40	49	63	23	69	466	0.19	K	-0.04	0.17	Option C replaced
16	50	47	44	96	60	28	69	52	20	466	0.13	K	0.15	0.07	Distractors are ok
17	63	43	60	40	60	43	40	98	19	466	0.09	0.09	0.07	K	Distractors are ok
18	40	108	60	40	57	50	66	25	20	466	K	0.09	0.03	0.18	Distractors are ok
19	50	94	67	50	57	33	49	44	22	466	K	0.07	0.1	0.02	Distractors are ok
20	40	120	60	38	57	41	66	24	20	466	K	0.09	0.07	0.18	Distractors are ok
21	60	36	47	107	49	33	67	47	20	466	0.1	K	0.07	0.09	Distractors are ok
22	60	57	63	38	46	98	54	32	18	466	0.01	0.12	K	0.09	Distractors are ok
23	50	120	66	40	60	36	57	37	0	466	K	0.11	0.1	0.09	A good item
24	52	25	56	100	64	36	61	82	-10	466	0.12	K	0.12	-0.09	Option D replaced
25	63	34	64	90	56	95	50	14	0	466	0.12	-0.11	K	0.15	Option B replaced
26	66	29	57	37	40	120	60	38	19	466	0.16	0.09	K	0.09	A good item
27	88	71	47	26	42	100	56	36	0	466	0.07	0.09	K	0.09	Distractors are ok



28	56	40	72	85	40	16	65	92	0	466	0.07	-0.06	0.1	K	Option B replaced
29	73	34	52	54	60	42	38	106	7	466	0.17	-0.01	0.02	K	Option B replaced
30	71	70	54	37	52	100	56	26	0	466		0.07	K	0.13	Distractors are ok
31	50	73	64	95	64	29	55	36	0	466	-0.1	K	0.15	0.08	Option A replaced
32	47	38	54	90	70	79	26	62	0	466	0.04	K	-0.04	-0.15	Option C&D replaced
33	63	88	45	37	56	36	69	72	0	466	K	0.03	0.09	-0.01	Option D replaced
34	66	50	43	110	65	38	59	35	0	466	0.07	K	0.12	0.1	A good item
35	90	74	50	39	52	95	41	25	0	466	0.07	0.05	K	0.07	Distractors are ok
36	80	85	80	70	50	58	23	20	0	466	K	0.04	-0.03	0.01	Option C replaced
37	66	40	37	33	60	61	70	99	0	466	0.11	0.02		K	Option C replaced
38	83	79	56	19	67	90	27	45	0	466	0.02	0.16	K	-0.08	Option D replaced
39	66	100	56	36	62	46	49	51	0	466	K	0.09	0.07	-0.01	Option D replaced
40	52	22	64	54	35	70	82	87	0	466	0.13	0.04	-0.15	K	Option C replaced
41	66	146	16	46	27	50	44	60	11	466	K	-0.13	-0.1	-0.07	All options (B, C&D) replaced
42	48	62	60	61	67	64	44	60	0	466	-0.06		K	-0.06	All options (B, C&D) replaced
43	102	118	46	38	39	48	30	45	0	466	K	0.03	-0.04	-0.06	Option C & D replaced
44	35	61	32	60	36	60	120	42	20	466	-0.11	-0.12	-0.1	K	Item amended
45	82	70	14	63	113	58	36	30	0	466	0.05	-0.21	K	0.03	Item amended
46	20	53	142	60	32	72	29	38	20	466	-0.14	0.35	K	-0.04	Item amended
47	54	39	41	62	72	65	81	58	-6	466	0.06	K	0.03	0.09	Distractors are ok
48	90	39	56	46	30	47	108	50	0	466	0.22	0.04	-0.07	K	Item amended
49	26	87	80	24	29	44	37	52	87	466	-0.26	0.24	-0.06	K	Option A & C replaced
50	80	85	80	70	50	58	23	20	0	466	K	0.04	-0.03	0.01	Option C replaced
51	66	40	70	99	37	33	61	37	23	466	0.11	K	0.02	0.1	A good item
52	83	79	56	19	67	90	45	27	0	466	0.02	0.12	K	0.08	Distractors are ok
53	66	100	56	36	62	46	49	51	0	466	K	0.09	0.07	-0.01	Option D replaced
54	52	22	82	87	35	70	64	54	0	466	0.13	K	-0.15	0.04	Option C replaced
55	50	10	70	36	80	53	24	127	16	466	0.17	0.15	0.12	K	A good item
56	55	17	68	48	40	118	60	30	30	466	0.16	0.09	K	0.13	A good item
57	20	132	74	08	60	56	69	27	20	466	K	0.28	0.02	018	A good item
58	60	32	48	108	61	28	74	45	10	466	0.12	K	014	0.12	A good item
59	52	30	63	36	34	128	64	35	24	466	0.09	0.12	K	0.12	A good item
60	64	30	60	30	63	23	36	140	20	466	0.15	0.13	017	K	A good item
61	50	141	61	37	44	23	58	22	30	466	K	0.1	0.09	0.15	A good item
62	46	110	52	34	65	52	60	37	10	466	K	0.08	0.06	0.1	Distractors are ok



63	60	44	61	42	51	102	51	38	17	466	0.07	0.08	K	0.06	Distractors are ok
64	49	97	60	41	57	38	57	51	16	466	K	0.08	0.08	0.03	Distractors are ok
65	29	23	41	120	60	40	94	40	19	466	0.03	K	0.09	0.23	Distractors are ok
66	40	120	60	38	57	41	66	24	20	466	K	0.09	0.07	0.18	Distractors are ok
67	60	36	47	107	67	47	49	33	20	466	0.1	K	0.09	0.07	Distractors are ok
68	60	57	46	98	54	32	63	38	18	466	0.01	K	0.09	0.12	Distractors are ok
69	60	40	57	50	66	25	40	108	20	466	0.09	0.03	0.18	K	Distractors are ok
70	67	50	50	94	49	44	57	33	22	466	0.07	K	0.02	0.1	Distractors are ok
71	57	20	60	22	76	21	30	58	122	466	0.16	0.16	0.24	K	A good item
72	64	10	58	161	28	06	81	40	18	466	0.23	K	0.09	0.18	A good item
73	70	20	50	10	43	176	60	23	14	466	0.21	0.17	-0.57	K	Option C replaced
74	63	42	41	107	61	38	58	26	30	466	0.09	-0.28	K	0.14	Item amended
75	28	160	60	06	26	10	92	47	37	466	-0.57	0.23	0.07	K	Item amended
76	36	136	60	20	46	30	71	43	24	466	-0.43	0.17	K	0.12	Item amended
77	23	124	50	10	70	36	80	53	20	466	-0.43	K	0.15	0.12	Item amended
78	55	17	68	48	40	118	60	30	30	466	0.16	0.09	K	0.13	A good item
79	74	08	20	132	60	56	69	27	20	466	0.28	-0.48	0.02	K	Option B replaced
80	60	32	61	28	75	45	48	108	9	466	0.12	0.14	0.13	K	A good item
81	52	30	63	38	34	128	64	35	22	466	0.09	0.12	-0.40	K	Item amended
82	64	30	60	30	36	140	63	23	20	466	K	0.13	-0.45	0.17	Item amended
83	44	23	50	141	58	22	61	37	30	466	0.09	-0.39	0.15	K	Item amended
84	70	25	40	119	50	43	63	36	20	466	0.19	-0.34	0.03	K	Item amended
85	50	06	60	49	63	23	50	116	49	466	0.19	0.05	K	-0.28	Item amended
86	50	47	60	28	44	96	69	52	20	466	0.01	K	-0.22	0.07	Item amended
87	63	43	60	40	40	98	60	43	19	466	0.09	K	-0.25	0.07	Item amended
88	60	40	40	108	57	50	66	25	20	466	0.09	-0.29	K	0.18	Item amended
89	67	50	50	94	57	33	49	44	22	466	0.07	-0.19	K	0.02	Item amended
90	46	16	66	146	50	27	60	44	11	466	0.13	-0.34	K	0.07	Item amended
91	64	67	62	48	61	60	60	44	0	466	-0.01	K		0.07	Item amended
92	38	46	48	39	45	30	102	118	0	466	K	0.04	0.06	-0.07	Option D replaced
93	61	35	42	120	60	32	60	36	20	466	K	-0.33	0.12	0.1	Option B replaced
94	82	70	63	14	58	113	30	36	0	466	K	0.21	-0.24	-0.03	Item amended
95	60	142	53	20	72	32	38	29	20	466	-0.35	0.14	K	0.04	Item amended
96	54	39	62	41	65	72	58	81	-6	466	0.06	K	-0.03	-0.1	Item amended
97	90	39	46	56	50	108	47	30	0	466	0.22	K	-0.25	0.07	Option C replaced



98	24	80	44	29	26	87	52	37	87	466	-0.24	0.06	K	0.06	Option A replaced
99	60	38	120	40	57	41	66	27	17	466	0.09	K	0.07	0.17	Item amended
100	60	36	47	107	49	33	67	47	20	466	0.1	K	0.07	0.09	Distractors are ok
101	60	59	46	98	63	38	54	32	16	466		K	0.11	0.09	A good item
102	66	40	60	36	57	37	50	120	0	466	0.11	0.13	0.09	K	A good item
103	56	100	52	25	64	36	82	61	-10	466	K	0.12	0.12	0.09	A good item
104	63	34	90	64	56	95	50	14	0	466	0.12	0.11	K	0.15	A good item
105	56	95	50	14	64	90	63	34	0	466	K	0.15	-0.11	0.12	Option C replaced
106	66	29	57	37	40	120	60	38	19	466	0.16	0.09	K	0.09	A good item
107	88	71	42	100	47	26	56	36	0	466	0.07	K	0.09	0.09	Distractors are ok
108	56	40	85	72	65	92	40	16	0	466	0.07	0.06	K	0.1	Distractors are ok
109	73	34	38	106	60	42	52	41	20	466	0.17	K	0.07	0.05	Distractors are ok
110	71	70	52	100	54	37	56	26	0	466		K	0.07	0.13	Distractors are ok
111	64	95	50	73	64	29	55	36	0	466	K	-0.1	0.15	0.08	Option B replaced
112	47	38	54	90	79	70	62	26	0	466	0.04	K	0.04	0.15	Distractors are ok
113	45	37	56	36	69	72	63	88	0	466	0.03	0.09	-0.01	K	Option C replaced
114	43	110	66	50	65	38	59	35	0	466	K	0.07	0.12	0.1	A good item
115	50	39	90	74	41	25	52	95	0	466	0.05	0.07	0.07	K	Distractors are ok
116	80	70	58	80	80	85	23	20	-30	466	0.04	-0.09	k	0.01	Option B replaced
117	46	110	65	52	60	37	52	34	10	466	-0.27	K	0.1	0.08	Item amended
118	60	44	61	42	51	36	51	102	19	466	0.07	0.08	0.06	k	Distractors are ok
119	60	41	49	97	57	38	57	51	16	466	0.08	k	0.08	0.03	Distractors are ok
120	29	23	94	40	40	120	60	40	20	466	0.03	0.23	k	0.09	A good item



Reliability of the PTMEAT

Estimate of PTMEAT reliability using test retest gave a coefficient index of 0.82.

The reliability estimates were established through test re-test method and gave a coefficient of 0.82. The test was found to be of good quality, valid and highly reliable. The PTMEAT is therefore recommended for use in assessing undergraduate students in measurements and evaluation.

Validation of PTMEAT

The results of Test Blueprint for Proficiency Test in Measurement and Evaluation Proficiency Test (PTMEAT) Item Analysis

Table 3: Test Blue Print for Proficiency Test in Measurement and Evaluation

Content	Know	Com	Appli	Interpre	Total
History of Testing	8	5	6	3	22
Types of Evaluation Instruments and Characteristics	4	6	4	1	15
Test Construction and Item Analysis	6	8	2	3	19
Test Administration, Scoring and Interpretation	6	10	3	1	20
Basic Statistics	-	8	12	4	24
Total	24	37	27	12	100

Abbreviations: Know=Knowledge; Com=Comprehension; Appli=Application; Interpre=Interpretation

Validity of the was done by matching the draft test items from the objectives (see table 1) and presented to three experts (one from Nnamdi Azikwe University, Awka and two from University of Maiduguri) in measurement and evaluation for item review. These experts guaranteed that the instrument had a strong and positive content and coverage validity in which each items represented the area of coverage which was been investigated. This implied that all the objectives and content areas were well covered in the blue print.

Discussion

The quality of test item is evident in the appropriateness and effectiveness of the test item parameters (difficulty, discrimination and distractor indices) obtained from this study, 100 test items made the requirement of good items with appropriate difficulty and discrimination indices that were retained. Test item numbers 2, 18, 21, 22, 26, 45, 51, 64, 67, 86, 88, 99, 101, 102, 105, and 109, had difficulty indices that was less than 0.30. this indicate that they were very difficult items. Test items numbers 5 and 9 were found to be very easy. Items number 32, 67 and 86 had negative discrimination indices. This revealed that low ability Undergraduates students' performance better on those items than high ability Undergraduate students. Such items discriminated but in the negatives direction. Test items number 2, 5, 8, 26, 29, 32, 45, 48, 63, 67, 69, 84, 86, 88, 95, 96, 103, 105, 111, 115 and 119 had low but positive discrimination values. This implied that undergraduate students who incorrectly answer this test items also scored high on the test overall, while students correctly answered the items scored low on the test overall.

The study indicated that in table 3 (distractor indices) that out of 100 PTMEAT items retained 1, 2, 3, 4, 5, 6, 7, 10, 13, 14, 23, 26, 34, 51, 55, 56, 57, 58, 59, 60, 61, 71, 72, 78, 80, 101, 102, 103, 104, 106, 114 and 120 have good items characters and some distractor indices of the 9, 15, 24, 25, 28, 29, 31, 32, 33, 36, 37, 38, 39, 40, 41, 42, 43, 49, 50, 53, 54, 73, 79, 92, 93, 97, 98, 105, 111, 113, and 116 were replaced. The positive values indicated that the distractor were good. It also revealed that more of the undergraduate students in the low ability group selected the distracter than those in the high achiever group. The negative value of the indices revealed that the distracters were weak. The distracter with negative values were reviewed then replaced or amended for improvement and ease of the test takers. Other test items were also improved by restricting the manner of questioning to reduce confusions in answering. This is in line with study conducted by Eleje et al (2016).

Validity of the PTMEAT was done by matching the draft test items from the objectives and presented to three experts (one from Nnamdi Azikwe University, Awka and University of Maiduguri) in measurement and evaluation for item review. These



experts guaranteed that the instrument had a strong and positive content and coverage validity in which each items represented the area of coverage which was been investigated. This implied that all the objectives and content areas were well coverage in the blue print. This the measurement and evaluation test items has a good content validity. The study agreed with Ofon U. J., et al. (2022) who studied, developed, validated, and standardized a statistical literacy test for university undergraduates in Cross River State, Nigeria. The result showed that only 25 items survived the last item assessment. The instrument was found to have strong construct validity as the items in the correlation matrix were all positively correlated. The statistical literacy scale showed convergence with statistical reasoning but diverged with statistical anxiety.

The reliability estimates that were obtained was 0.821. The coefficient was high indicating that the instrument consistently measured the traits designed to measure over a period of time. According to Ceniza and Cereno (2012) in Eleje, Abanobi and Obasi (2017) the reliability coefficient within the range of 0.81 to 1.0 signified high reliability.

Conclusion

A good, valid and reliable measurements and evaluation proficiency test for Universities Undergraduate students was developed in this study. This can be at least based on the results of the reliability and item analysis of the test conducted. The test analysis conducted on the test items revealed that the test overall difficulty was within the range of 0.30 to 0.70 and the discrimination indices were within the range of 0.20 to 0.50. This implies that PTMEAT has a moderate level of difficulty and the discrimination indices. This implied that the currently developed measurement and evaluation proficiency test PTMEAT is highly reliable.

Therefore, it could be concluded that the measurement and evaluation proficiency test developed in this study is of good quality, valid and highly reliable PTMEAT can be used in assessing Universities Undergraduate student's proficiencies in measurements and evaluation. It is an instrument that can measure the desired trait of Undergraduate students in the Universities.

Recommendation

The fact that the findings of this study revealed that the PTMEAT is valid and of good quality, the researchers recommended that the test be used by lecturers to assess their students. similar test items should be developed for other education courses in the university

References

- AERA, APA, & NCME. (2014). Standards for educational and psychological testing. American Educational Research Association.
- Alderson, J. C. (2005). Diagnosing foreign language proficiency: The interface between learning and assessment. Continuum.
- Baykal A (2011). Measurement and evaluation in education. Ankara, Turkey: Pegem Academy Publishing
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>
- Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice*, 30(1), 3–12. <https://doi.org/10.1111/j.1745-3992.2010.00195.x>
- Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. Pearson Education.
- Brown, H. D., & Abeywickrama, P. (2019). *Language assessment: Principles and classroom practices* (3rd ed.). Pearson.
- Ceniza MJJ, Cereno JL (2012). Test construction and validation procedures. *International Journal of Educational Measurement*, 5(2), 45–58.
- Eleje LI, Abanobi C, Obasi C (2017). Development and validation of a mathematics achievement test for secondary school students. *Journal of Educational Research and Evaluation*, 16(1), 23–35.
- Eleje LI, Esomonu NPM, Abanobi CC (2016). Item analysis of teacher-made tests using classical test theory. *Journal of Educational Measurement and Evaluation*, 15(2), 101–115.



- Esomonu NPM (2021). Instrumentation model for test development and validation. Awka, Nigeria: Nnamdi Azikiwe University Press.
- Esomonu NPM, Agbonkpolo MU (2010). Teachers' competence in test construction and validation in Nigerian secondary schools. *Nigerian Journal of Educational Measurement and Evaluation*, 10(1), 55–66.
- Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15(3), 309–334. https://doi.org/10.1207/S15324818AME1503_5
- Linn, R. L., & Gronlund, N. E. (2000). *Measurement and assessment in teaching* (8th ed.). Prentice Hall.
- Mehta G, Mokhasi V (2014). Item analysis of multiple-choice questions—An assessment of the assessment tool. *International Journal of Health Sciences and Research*, 4(7), 197–202.
- Ministry of National Education (2017b). Teacher strategy document. Ankara, Turkey: Ministry of National Education.
- Ministry of National Education (2019). Education vision 2023. Ankara, Turkey: Ministry of National Education.
- Nitko AJ (2001). *Educational assessment of students* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.
- Nitko, A. J., & Brookhart, S. M. (2014). *Educational assessment of students* (7th ed.). Pearson.
- Oduor JA, Kibul JR (2018). Development and validation of a measurement and evaluation proficiency scale for undergraduate students. *African Journal of Educational Research*, 12(3), 89–102.
- Ofon UJ, Akpan AA, Essien EE (2022). Development, validation and standardization of a statistical literacy test for university undergraduates in Cross River State, Nigeria. *Journal of Educational Measurement and Evaluation*, 21(2), 134–156.
- Popham, W. J. (2017). *Classroom assessment: What teachers need to know* (8th ed.). Pearson.