



Item Analysis of Objective Structured Practical Examination (OSPE) Used as an Assessment Tool for First-Year Microbiology Students.

Sylvester Chibueze Izah^{1*}, Tamaraukepreye Catherine Odubo¹, Victor Emeka Ajumobi¹ and Olugbenro Osinowo²

¹Department of Microbiology, Faculty of Science, Bayelsa Medical University, Yenagoa, Bayelsa State, Nigeria.

²Department of Surgery, Faculty of Clinical Sciences, Bayelsa Medical University, Yenagoa, Bayelsa State, Nigeria.

ARTICLE INFO

Article No.: 112621137

Type: Research

Full Text: [PDF](#), [HTML](#), [PHP](#), [EPUB](#)

Accepted: 29/11/2021

Published: 31/03/2022

*Corresponding Author

Sylvester Izah

E-mail: chivestizah@gmail.com

Phone: +2347030192466

Keywords: Assessment tool, Internal Consistency, Item analysis, Microbiology items, OSPE.

ABSTRACT

Item analysis is used to examine students' responses to items to determine the quality of an assessment tool. This study aimed at assessing the quality of the objective structured practical examination (OSPE) on Introductory Microbiology. 41 first-year Microbiology major students of Bayelsa Medical University, Nigeria took part in the OSPE containing 40 items. Marks were not deducted for wrong answers as decided by the department and each item carried one mark, and 40% was the pass mark. The items were analyzed for the difficulty index, discrimination index, distractor efficiency, and reliability index (Cronbach's alpha). Also, key distributions, numbers, and percentage passed were determined. Results showed that the discrimination index had 15 (37.5 %), 2 (5.0 %), and 3 (7.5 %) translated as excellent, good, and acceptable. The difficulty index revealed that 4 (10.0 %) of the items were ideal, while the remaining 36 (90.0 %) were difficult. The 40 items had 160 distractors, of which 71 (44.4 %) and 89 (55.6 %) were functional and non-functional distractors, respectively. The difficulty index indicated positive significant relationship with the discrimination index ($r = 1.000$) and distractor efficiency ($r = 0.408$) at $p = 0.01$. The overall reliability analysis of the item was 0.754, an indication that it is good for classroom assessment, but some items need improvement. Too easy items and poor distractors may have caused the poor difficulty index. Therefore, there is the need for item flaws and technical pitfalls to be carried out to correct the errors in subsequent assessments. From the findings in this study it is recommended that OSPE be more widely adopted among more science-based departments and item analysis be a standard practice in departments of every university.

1.0 BACKGROUND

Assessment is a very vital key in the education process. It is crucial in determining the nature of the learning person, their way of knowing, what the learner knows, need to know, and what learning and mediating processes are associated with effective teaching and learning for each learner (CERI, 2008, Armour-Thomas and Gordon, 2013). In recent years, educational research and policies have increasingly focused on standardized forms of learning-centered assessments to improve knowledge, skills, and disposition for living in a competitive global society. An assessment provides feedback about the strengths and weaknesses of the students' performances about a given task which can inform subsequent decisions about curriculum and instruction (Armour-Thomas and Gordon, 2013). In addition, the assessment tools also need to be evaluated because the assessment method can influence students learning (Vishwakarma, 2016).

A significant part of a microbiology curriculum is an appropriate assessment of the student's practical capabilities in the laboratory. For general student assessments, different types of examinations are used, such as Multiple choice examinations (MCQ), short answer examination (SAQ), and theory (Essay) examinations, all of which are vital for only assessing the knowledge, i.e., cognitive domain of the students (Relwani et al., 2016). On the other hand, practical examinations are essential for determining the cognitive, psychomotor, and affective domains, and thus an effective system of evaluation should be applied (Relwani et al., 2016). Assessment drives learning; therefore, adopting a suitable assessment tool would ensure constructive alignment between goals and learning outcomes (Bhat et al., 2020).

In the clinical sciences, the examination of practical skills and competence is critical for proper medical education. Still, conventional clinical and practical examinations had several limitations, especially in terms of their outcome. The traditional practical assessments lacked objectivity (Frantz et al., 2013). Although grading should depend on the student's competence variability in the experimental process, standardization and examiners' decision were also concerns (Jaswal et al., 2015). These defects in traditional practical examinations, especially in clinical sciences, gave rise to the development of new examination systems which can test all the objectives (Munjal et al., 2011).

The objective structured practical examination (OSPE) was modified from Objective Structured Clinical Examination (OSCE) in the 1970s (Harden and Cairncross, 1980; Frantz et al., 2013; Mard and Ghafouri, 2020), and has gained worldwide acceptance as a method for clinical skills assessment because of its uniformity, reliability, validity, and practicability (Nigam and Mahawar, 2011), standardization, and absence of variability in an experiment, scores, etc. In addition, it

provides a more objective method of assessment, covers a broad scope, maximizes reliability, allows individual students to display a full range of their knowledge, skills, and abilities, which are evaluated in a comprehensive, structured and consistent manner and provides a measurement of skills (Mokkapati et al., 2016, Vijaya and Alan, 2014, Frantz et al., 2013).

The OSPE is a suitable examination system comprising a series of stations where students work through different tasks to test various skills applying to all learning domains (Munjal et al., 2011). All students are expected at all stations and spend equal time at each station, eliminating group work by students. All the students take the same examination, which is standardized.

The use of OSPE has been reported in many Universities with great benefits (Vishwakarma et al., 2016). OSPE maximizes reliability in assessment and allows students to display the full range of their knowledge, skills, and abilities. OSPE is very popular in the medical sciences. Information about OSPE in non-health sciences is very scanty, especially in developing countries like Nigeria. The Department of Microbiology Bayelsa Medical University, Bayelsa State, Nigeria, adopted OSPE as a tool for assessing students in their practical examination because of its merits over the traditional practical assessments. The practical assessment of introductory microbiology, a first-year microbiology course for microbiology major students was carried out using OSPE. Since this is the first time an OSPE is being conducted in the institution, item analysis should be carried out to ascertain the reliability of the examination. Hence, the focus of this paper is to carry out an item analysis of OSPE as an assessment tool for first-year microbiology major students. The findings of this study will be helpful to the scientific community, especially the Department of Microbiology, in decision-making about effective assessment methods for first-year Microbiology students.

2.0 METHODS

Design techniques and organizational structure of OSPE in Bayelsa Medical University, Nigeria

Before the OSPE, the course lecturers informed the students of the processes that would be carried out for the practical examination. The exercise was carried out by the course lecturers (examiners) with the assistance of eight Laboratory Technologists and a Professor in the University who have carried out OSPE/OSCE before now.

A total of forty-one first-year microbiology major students participated in the study. The students were divided into two batches. Three laboratories were used for the exercise; two of the three were arranged with twelve stations each. The third laboratory was used as a quarantine room for the students who belonged to the

second batch. The questions were pasted on each of the stations before students entered the examination hall or laboratories. Twenty-four students belonging to the first batch were moved into the two designated laboratories, twelve students in each examination room. Hence there were twelve stations (including ten workstations and two rest stations). The rest stations were arranged at stations 6 and 12. The students were placed in such a way that each student in a station spends only five minutes. The duration spent for each workstation was regulated by a laboratory technologist who served as the timekeeper during the OSPE. At the end of the exercise for the first batch, the answer scripts were retrieved, and they were moved to the third laboratory, which served as the quarantine room. At the same time, the students for the second batch were moved to the two laboratories used for the OSPE itself. The movement was carried out in a way that both groups of students did not meet. Twelve students were placed in a laboratory from the second batch, and the remaining five students were placed in the other laboratory. The laboratory with only five students was invigilated by four laboratory technologists and supervised by one examiner.

Structure of OSPE Questions

The OSPE consisted of forty single response stems, four alternatives (distractors), and one key (correct answer). Each correct response was awarded one mark, and no marks were awarded for blank or incorrect answers. Negative marks were not awarded for wrong answers. The maximum possible score in the examination was forty. The scores of all students were arranged in the order of merit and divided into three groups. The upper one-third students (H) (27%) will be considered to have high ability, and the lower one-third (L) (27%) will be deemed to be lower ability (Hingorjo and Jaleel, 2012). Out of the 41 students, 11 will be in the high group and eleven in the lower group, while nineteen will be in the middle group and will not be used in this study. Based on the data, the discrimination index (DIS), difficulty index (DI) or (p-value), and distractor efficiency (DE) were calculated following the method previously described by Hingorjo and Jaleel (2012), Gajjar et al. (2014).

$$\text{p-value or difficulty index (DI)} = [(H+L)/N] \times 100 \text{ -----[1]}$$

$$\text{Discrimination index (DIS)} = 2 \times [(H-L)/N] \text{-----[2]}$$

Where N is the total number of students in both high and low groups, and H and L are the correct responses in high and low groups, respectively.

Items with $P < 30\%$ are considered difficult, $p=30-70\%$ are ideal, and $P > 70\%$ are considered easy. Discrimination index (DIS) is the ability of an item to differentiate between students of higher and lower abilities, and it ranges between 0 and 1.

D=Negative (Defective item/wrong key), $D=0-0.19$ (Poor discrimination), $D=0.2-0.29$ (Acceptable discrimination), $D=0.3-0.39$ (Good discrimination) and $D > 0.4$ (Excellent discrimination). An item contains a stem, five options, including one correct option (Key), and four incorrect alternatives (Distractors). Non-functional distractors (NFD) were options other than the key selected by $< 5\%$ of students, and functional distractors were the option (distractors) selected by 5% or more students. DE ranged from 0% to 100% and was determined based on the number of NFDs in an item. If an item contains four, three, two, one, and zero NFDs, the DE would be 0%, 25%, 50%, 75%, and 100%, respectively.

Statistical analysis

SPSS version 20 and Microsoft Excel were used to carry out the statistical analysis. Descriptive statistics (mean, mode, median, and standard deviation), Pearson's correlation, Chi-square, analysis of variance, and reliability (Cronbach's alpha) were carried out at varying levels. Charts were used to show the total score using histograms with binomial distribution curves, bar charts for key distribution and percentages of students that passed, and interpolation lines for criteria for the different indices (DIS, DI, and DE).

RESULTS

Table 1 shows the discrimination and difficulty indices of OSPE among first-year Microbiology major students. The DI and DIS ranged from 0.00 – 50.00 with a mean $\pm$ standard deviation of 14.66 ± 12.79 and 0.00 – 100.00 with mean $\pm$ standard deviation of 0.29 ± 0.26 , respectively. However, the DIS and DI mean value was within acceptable and very difficult ranges, respectively. The percentage distribution of the DIS and DI criteria are shown in Figures 1 and 2, respectively. Out of the 40 items, 15 representing 37.50%, 2 representing 5.00%, 3 representing 7.50%, and 20 representing 50.00% showed excellent, good, acceptable, and poor discrimination, respectively (Figure 1). These frequencies showed significant variations ($\chi^2 = 58.72$, $P=0.000$). For the DI, 4 representing 10.00% and 36 representing 90.00% showed that the items were ideal and difficult, respectively (Figure 2). Statistically, there were variations ($\chi^2 = 64.00$, $P=0.000$) between the two criteria that the values of this study fall within.

Table 1: Discrimination and difficulty indices of OSPE among first-year Microbiology major students

Items	Difficulty index	Criteria	Discrimination index	Criteria
1.	50.00	Ideal	1.00	Excellent
2.	13.64	Difficult	0.27	Acceptable
3.	0.00	Difficult	0.00	Poor
4.	4.55	Difficult	0.09	Poor
5.	22.73	Difficult	0.45	Excellent
6.	22.73	Difficult	0.45	Excellent
7.	0.00	Difficult	0.00	Poor
8.	0.00	Difficult	0.00	Poor
9.	0.00	Difficult	0.00	Poor
10.	0.00	Difficult	0.00	Poor
11.	22.73	Difficult	0.45	Excellent
12.	13.64	Difficult	0.27	Acceptable
13.	9.09	Difficult	0.18	Poor
14.	9.09	Difficult	0.18	Poor
15.	4.55	Difficult	0.09	Poor
16.	13.64	Difficult	0.27	Acceptable
17.	22.73	Difficult	0.45	Excellent
18.	22.73	Difficult	0.45	Excellent
19.	4.55	Difficult	0.09	Poor
20.	22.73	Difficult	0.45	Excellent
21.	0.00	Difficult	0.00	Poor
22.	9.09	Difficult	0.18	Poor
23.	4.55	Difficult	0.09	Poor
24.	4.55	Difficult	0.09	Poor
25.	27.27	Difficult	0.55	Excellent
26.	36.36	Ideal	0.73	Excellent
27.	9.09	Difficult	0.18	Poor
28.	4.55	Difficult	0.09	Poor
29.	4.55	Difficult	0.09	Poor
30.	4.55	Difficult	0.09	Poor
31.	27.27	Difficult	0.55	Excellent
32.	18.18	Difficult	0.36	Good
33.	27.27	Difficult	0.55	Excellent
34.	9.09	Difficult	0.18	Poor
35.	40.91	Ideal	0.82	Excellent
36.	22.73	Difficult	0.45	Excellent
37.	22.73	Difficult	0.45	Excellent
38.	36.36	Ideal	0.73	Excellent
39.	18.18	Difficult	0.36	Good
40.	0.00	Difficult	0.00	Poor
Mean	14.66	Difficult	0.29	Acceptable
Standard deviation($\pm$)	12.79	-	0.26	-
Minimum	0.00	-	0.00	-
Maximum	50.00	-	1.00	-

Source: Authors

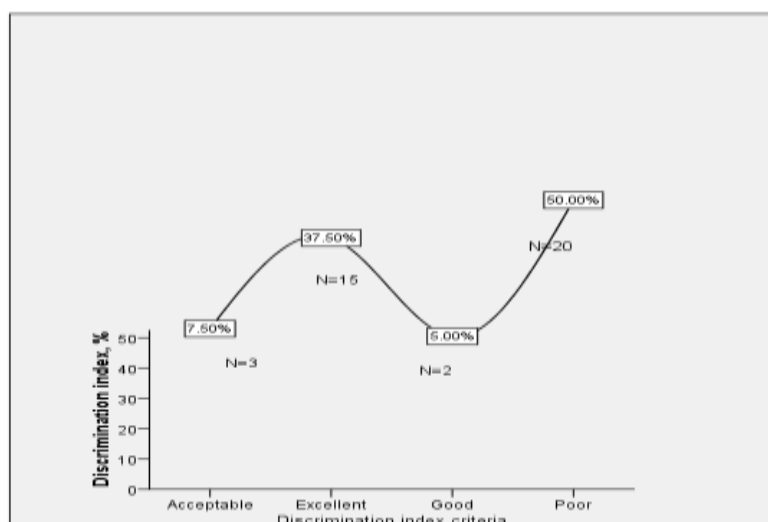


Figure 1: Percentage distribution of the discrimination index criteria.

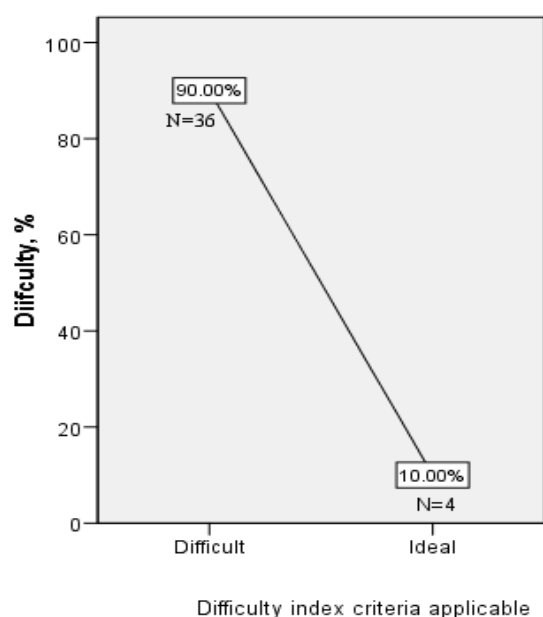


Figure 2: Percentage distribution of the difficulty index criteria

Table 2 shows the distractor efficiency and implication of the outcome of OSPE among first-year students of the Microbiology programme. The DE ranged from 0.00 – 100.00%, with a mean of $44.74 \pm 30.74\%$. Out of 160 distractors, 71 (44.38%) and 89 (55.62%) were functional and non-functional distractors respectively (Table 3). Thus, the majority of the items need one or more distractors revised. The percentage distribution of the distractors is shown in Figure 3. Among the 40 items, 3 (7.50%), 8 (20.00%), 15 (37.50%), 5 (12.50%) and 9 (22.50%) had 0, 1, 2, 3 and 4 non-functional distractors, respectively. These frequencies showed significant variations ($\chi^2 = 25.75$, $P=0.000$).

Table 4 shows Pearson's correlation between the DI, DIS, and DE. The indices (DI, DIS, and DE) showed very strong positive significant relationships at $p=0.01$.

The key distribution of OSPE among first-year microbiology major students is shown in Figure 4. The keys had equal (8 each 20.00%) distribution.

The distribution of the OSPE score according to the grading criteria is shown in Figure 5. Out of the 41 students that participated in the OSPE 18, 11, 9, 2 and 1 representing 43.90%, 16.80%, 22.00%, 4.90% and 2.40% scored A (70 – 85%), B (60 – 69%), C (50 – 59%), D (45 – 49%) and E (40 – 44%), respectively. Statistically, there was variation ($\chi^2 = 58.90$, $P=0.000$). The distribution (histogram showing the normal curve) of the students' scores is shown in Figure 6. The histogram showed skewness $\pm$ standard error of skewness of -0.248 ± 0.369 . The mean and standard deviation was 26.46 ± 4.70 . The least score was 16, which is the pass mark.

Table 2: Distractor efficiency and options distribution (4 distractors + 1 key) of OSPE among first-year students of the Microbiology programme

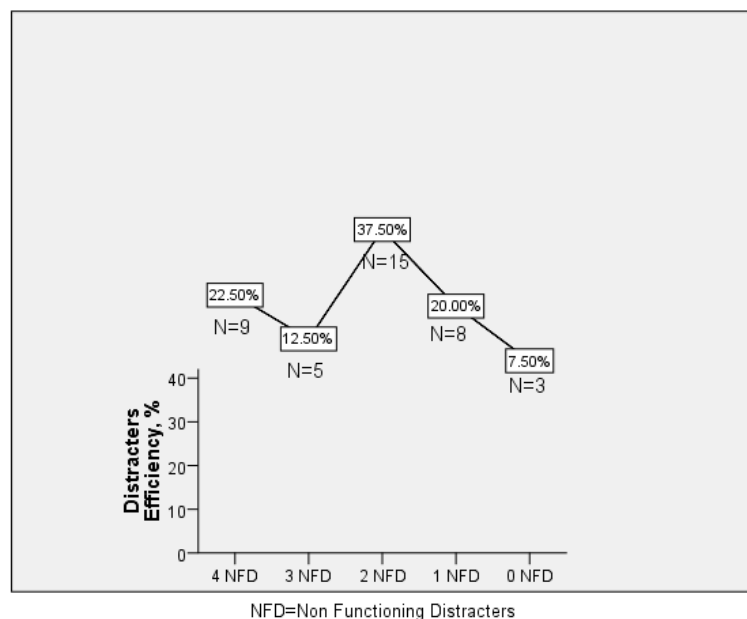
Items	Options (4 distractors + 1 key)					Distractor efficiency, %
	A	B	C	D	E	
1	2(4.90)	2(4.90)	29(70.70)	4(9.80)	4 (9.80)	50
2	6(14.60)	16(39.00)	3(7.30)	15(36.60)	1(2.40)	75
3	8(19.50)	10(24.40)	13(31.70)	7(17.10)	3(7.30)	100
4	0(0.00)	1(2.40)	0(0.00)	0(0.00)	40(97.60)	0
5	0(0.00)	2(4.90)	33(80.50)	6(14.60)	0(0.00)	25
6	5(12.20)	30(73.20)	2(4.90)	1(2.40)	3(7.30)	50
7	100(100.00)	0(0.00)	0(0.00)	0(0.00)	0(0.00)	0
8	1(2.40)	8(19.50)	15(36.60)	1(2.40)	16(39.00)	50
9	100(100.00)	0(0.00)	0(0.00)	0(0.00)	0(0.00)	0
10	0(0.00)	0(0.00)	1(2.40)	39(95.10)	1(2.40)	0
11	4(9.80)	1(2.40)	2(4.90)	28(68.30)	6(14.60)	50
12	0(0.00)	9(22.00)	3(7.30)	8(19.5)	21(51.20)	75
13	5(12.20)	1(2.40)	29(70.70)	5(12.20)	1(2.40)	50
14	25(61.00)	3(7.30)	8(19.50)	3(7.30)	2(4.90)	75
15	0(0.00)	38(92.70)	1(2.40)	2(4.90)	0(0.00)	0
16	18(43.90)	15(36.60)	2(4.90)	1(2.40)	5(12.20)	50
17	1(2.40)	9(22.00)	9(22.00)	5(12.20)	17(41.50)	75
18	0(0.00)	2(4.90)	8(19.50)	26(63.40)	5(12.20)	50
19	8(19.50)	12(29.30)	0(0.00)	15(36.60)	6(14.60)	75
20	1(2.40)	22(53.70)	5(12.20)	6(14.60)	7(17.10)	75
21	0(0.00)	0(0.00)	41(100.00)	0(0.00)	0(0.00)	0
22	35(85.40)	2(4.90)	3(7.30)	1(2.40)	0(0.00)	25
23	1(2.40)	0(0.00)	40(97.60)	0(0.00)	0(0.00)	0
24	0(0.00)	3(7.30)	0(0.00)	5(12.20)	33(80.50)	50
25	6(14.60)	26(63.40)	0(0.00)	4(9.80)	5(12.20)	75
26	25(61.00)	4(9.80)	3(7.30)	4(9.80)	5(12.20)	100
27	0(0.00)	2(4.90)	0(0.00)	38(92.70)	1(2.40)	0
28	0(0.00)	0(0.00)	40(97.60)	1(2.40)	0(0.00)	0
29	2(4.90)	3(7.30)	1(2.40)	0(0.00)	35(85.40)	25
30	6(14.60)	2(4.90)	1(2.40)	0(0.00)	32(78.00)	25
31	3(7.30)	1(2.40)	11(26.80)	26(63.40)	0(0.00)	50
32	19(46.30)	10(24.40)	4(9.80)	3(7.30)	5(12.20)	100
33	2(4.90)	0(0.00)	27(65.90)	10(24.40)	2(4.90)	25
34	2(4.90)	13(31.70)	23(56.10)	3(7.30)	0(0.00)	50
35	26(63.40)	1(2.40)	6(14.60)	8(19.50)	0(0.00)	50
36	15(36.60)	1(2.40)	19(46.30)	4(9.80)	2(4.90)	50
37	28(68.30)	1(2.40)	7(17.10)	1(2.40)	4(9.80)	50
38	5(12.20)	24(58.50)	1(2.40)	2(4.90)	9(22.00)	50
39	4(9.80)	26(63.40)	4(9.80)	0(0.00)	7(17.10)	75
40	6(14.60)	0(0.00)	5(12.20)	30(73.2)	0(0.00)	50
Mean	-	-	-	-	-	44.38
Standard error	-	-	-	-	-	30.74
Minimum	-	-	-	-	-	0.00
Maximum	-	-	-	-	-	100.00

Source: Authors

Table 3: Summary of Distractor efficiency of OSPE among first-year students of Microbiology programme

Parameters	N	%
Total OSPE	40	-
Distractors (Total)	160	-
Functional Distractors	71	44.38
Non-functional distractors	89	55.62

Source: Authors

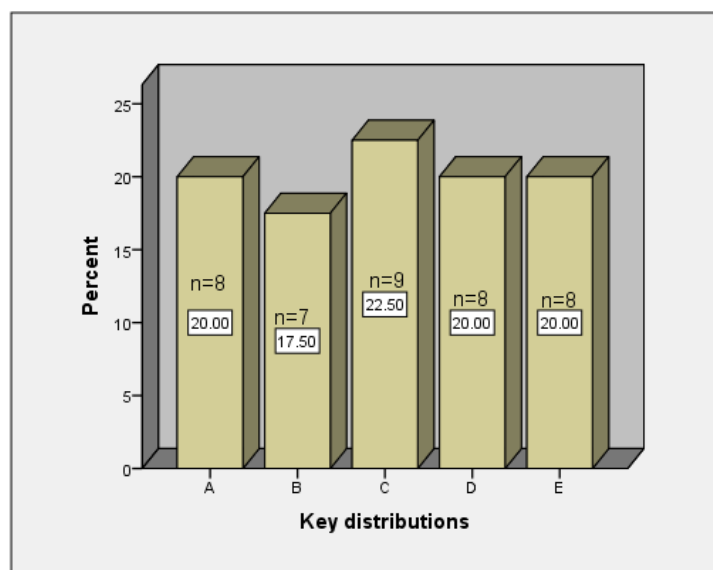
**Figure 3: Percentage distribution of the distractor efficiency criteria****Table 4: Pearson's correlation between difficulty index, discrimination index, and distractor efficiency**

Parameters	Difficulty index	Discrimination index	Distractor efficiency
Difficulty index	1		
Discrimination index	1.000**	1	
Distractor efficiency	0.408**	0.408**	1

**. Correlation is significant at the 0.01 level (2-tailed).

Source: Authors

N=40

**Figure 4: Percentage distribution of the keys**

The Cronbach's Alpha of all the OSPE items is 0.790. When some items are deleted, the internal consistency of the items (Cronbach's Alpha) decreases but is not

less than 0.766 (Table 5). The analysis of variance between the items showed that $p=0.000$ (Table 6).

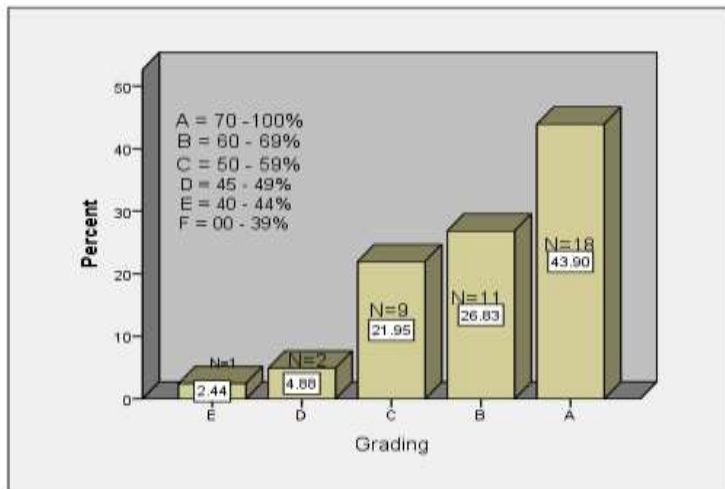


Figure 5: Distribution of OSPE scores according to the grading criteria

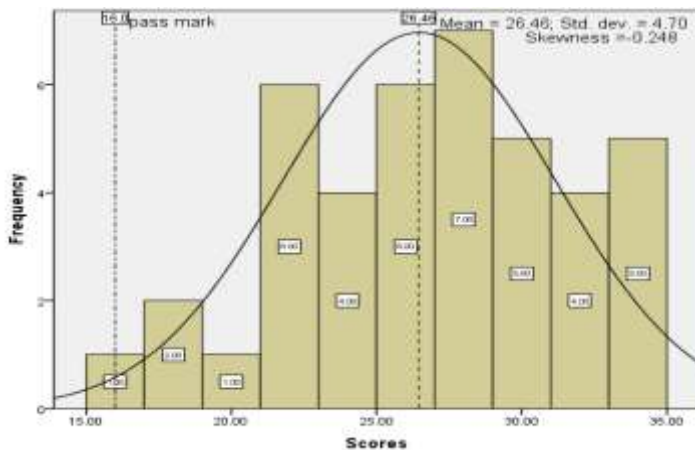


Figure 6: Distribution of OSPE total scores

Table 5: Cronbach's Alpha if each item is deleted
Item-Total Statistics

Items	Scale Mean if Item is Deleted	Scale Variance if Item is Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item is Deleted
Q1	27.2667	26.638	.604	.778
Q2	27.2000	28.314	.000	.791
Q3	27.8000	28.600	-.100	.802
Q4	27.2000	28.314	.000	.791
Q5	27.4667	25.838	.487	.776
Q6	27.3333	26.667	.419	.781
Q7	27.2000	28.314	.000	.791
Q8	27.9333	29.495	-.280	.807
Q9	27.2000	28.314	.000	.791
Q10	27.3333	26.952	.339	.783
Q11	27.4667	25.981	.455	.778
Q12	27.6000	25.543	.491	.775
Q13	27.6000	27.543	.097	.794
Q14	27.8000	27.457	.113	.793
Q15	27.2667	28.210	.014	.792
Q16	27.8000	28.886	-.152	.804
Q17	27.7333	25.781	.432	.778
Q18	27.4000	25.257	.693	.769
Q19	27.9333	28.924	-.166	.803
Q20	27.4667	26.695	.298	.784
Q21	27.2000	28.314	.000	.791
Q22	27.4000	27.686	.105	.792
Q23	27.2000	28.314	.000	.791
Q24	27.4667	26.410	.360	.782
Q25	27.6000	24.686	.669	.767
Q26	27.5333	26.838	.245	.787
Q27	27.2667	26.638	.604	.778
Q28	27.2000	28.314	.000	.791
Q29	27.3333	27.095	.299	.785
Q30	27.4000	27.400	.171	.789
Q31	27.6667	25.524	.484	.775
Q32	27.7333	25.781	.432	.778
Q33	27.5333	25.695	.481	.776
Q34	27.6000	29.114	-.193	.806
Q35	27.6000	24.400	.730	.764
Q36	27.4667	25.981	.455	.778
Q37	27.5333	25.838	.451	.777
Q38	27.4000	24.971	.766	.766
Q39	27.4667	27.410	.145	.791
Q40	28.2000	28.314	.000	.791

Table 6: Overall analysis of variance between the OSPE items

	Sum of Squares	df	Mean Square	F	Sig
Between People	9.910	14	.708	5.817	.000
Between Items	33.718	39	.865		
Within People Residual	81.157	546	.149		
Total	114.875	585	.196		
Total	124.785	599	.208		

Grand Mean = .7050

Source: Authors

DISCUSSION

Item analysis is usually carried out to examine students' responses to each item (OSPE). Its major focus is to determine the quality of the OSPE items and the overall test (examination). The discrimination index, which is usually used to show the high and low ability students, showed that 50.00% of the items had poor discrimination, while the DI is very high. Also, the non-functional distractors account for 55.62% of the total distractors. The DI, DIS, and DE suggest a need to improve the items because a weak design (especially for the distractors) may have contributed to the overall DI. Thus, these criteria are not providing adequate assessment (Izah et al., 2021).

In this study, item difficulty is seen in terms of frequencies with which those taking the OSPE chose the single best response instead of the distractors are associated with intrinsic characteristics of the items. It could also be seen that the items were too easy and answered correctly by nearly all the students. In contrast, the very difficult questions were responded to wrongly by both students (students with high and low abilities). This could be seen in 7 (17.50%) of the items. According to Chhaya et al. (2018), items answered correctly or wrongly by both groups of students should be removed. The mean difficulty index recorded in this study is lower than the values of 39.40% (Gajjar et al., 2014), 50.16% (Rao et al., 2017), 55.90% (Patel, 2017), 58.74% (Mahjabeen et al., 2017) and 57.62% (Chhaya et al., 2018) from MCQ examinations. The variation may be associated with intrinsic factors in the OSPE design.

The discrimination index ranges from 0 to 1.0. The findings of this study showed that 50.00% of the items had a poor discrimination index. According to Charanja et al. (2015), poor discrimination index is caused by ambiguous items, wrong keys, many correct answers, too easy or difficult items, and failure of teaching and learning sessions. From these criteria, the poor discrimination observed in 50.00% of items may be associated with the fact that the items were too easy, as shown in Figure 5, which showed that 38 students (92.70%) of the students scored $\geq 50.00\%$ of the total score and the remaining three students (7.30%) scored $\geq 40.00\%$ to < 50 of the total score.

Even though the mean DIS is within the acceptable range, the values were within previously reported ranges of 0.29 (Patel, 2017), 0.22 (Charania et al., 2015), but lower than the values of 0.34 (Rao et al., 2017) and 0.35 (Mahjabeen et al., 2017) recorded in some MCQ examinations.

The DE showed that approximately 44.00% of the distractors were effective, while the non-functional distractors accounted for over 55.00% of the total distractors. However, the findings of the non-functional distractors are related to the results of previous studies that have values of 15.00% (Patel, 2017), 5.00% (Rao et al., 2017), 11.4 (Gajjar et al., 2014), and lower than the value of 28.00% reported by Mahjabeen et al. (2017) in

some MCQ examinations. However, the non-functional distractors also influenced the DI and DIS. As seen in Table 4, there was a strong significant correlation between the three indicators (DE, DI, and DIS). The findings also agree with the work of Rao et al. (2017). According to Rao et al. (2017), reducing the number of distractors increases DIS and reliability level.

The study further revealed that no students failed the examination, but the distribution did not follow the standard binomial distribution curve. The skewness, which was less than -1, indicates the distribution is highly asymmetrical. However, the mean and standard deviation values provide estimates of the actual parameters of the curve. Thus, it gives information about the symmetrical distribution of the number of students that passed. But in this study, the skewness (degree of asymmetry or not symmetrical) is negative, having a longer tail in the left of the central maximum with mean values less than the mode and median values, an indication that most of the students are high scorers.

The reliability (internal consistency) is within 0.70 – 0.80 and 0.80 – 0.90, which are classified as suitable for a classroom assessment with few items requiring improvement and very good for classroom assessment (Patel, 2017). However, a value of 0.790 recorded in this study indicates that the overall test is still good and reliable. However, deletion of some of the items reduces the Cronbach's alpha value, which suggests that such items strongly influenced the reliability of the test. The Cronbach's Alpha value recorded in this study is slightly higher than the value of 0.702 previously reported by Patel (2017) in an MCQ examination. Again, the analysis of variance showed that $p=0.000$, an indication that the reliability of the items is quite different statistically.

CONCLUSIONS

Item analysis as an evaluation tool is beneficial to both students and teachers. This study assessed the quality of OSPE served to first-year microbiology major students, and the results showed that the mean DIS was within the acceptable range. However, 17.50% of the items need to be removed because of their poor discrimination power. The DI showed that 90% of the items are difficult, and 10% are ideal. None of the students scored below the 40% pass mark, indicating that most of the items were too easy. The DE showed that over 55.00% of the distractors were not effective. The Cronbach's Alpha showed that the items are classified as good for an assessment though improvement is required. Based on the DI and DE, there is a need to carry out item flaws and technical pitfalls analysis to improve the assessment scores, identify the difficult items, discriminate among the students, and remove or revise the non-functioning distractors. From the findings in this study it is recommended that this exercise be repeated for the next four years and the results compared to ensure assessments items are

standard. It is also recommended that OSPE be more widely adopted among more science-based departments and item analysis be a standard practice in departments of every university.

Acknowledgments

The authors would like to express their thanks to the following Laboratory Technologists of Bayelsa Medical University that participated in the invigilation of the OSPE; Ms. Timipre Grace Tuaboboh, Ms. Biembebe Virtuous Temple, Ms. Sebhaziba Benjamin Ezem, Ms. Blessing Muji Olagoke, Ms. Ann Tugwell Ototo, Mrs Christy Koroye, Mr Henry Ebiowei Alpha and Mr Samuel Philemon Bokene.

Ethical approval

Ethical approved was obtained from the Research and Ethics Committee of Bayelsa Medical University, Yenagoa, Bayelsa State, Nigeria with Ethical approval number REC/2021/0009.

Competing interests

The authors declare that they have no competing interests.

REFERENCES

1. Armour-Thomas, E., and E.W. Gordon, 2013. "Toward an understanding of assessment as a dynamic component of pedagogy". Gordon Commission on the future of assessment in education. Retrieved June 30 2020. https://www.ets.org/Media/Research/pdf/armour_thomas_gordon_understanding_assessment.pdf.
2. Bhat, D., P. Murugesh, and NB. Pushpa, 2020. "Objective structured practical examination: As an assessment tool in newly introduced competency based anatomy curriculum". *Indian Journal of Clinical Anatomy and Physiology*, 7(1): 81-86.
3. Centre for Educational Research and Innovation (CERI), 2008. "Assessment for learning formative assessment". Retrieved from oecd.org/site/educeri21st/40600533.pdf. Accessed December 16, 2020.
4. Charania, JS 2015. "Item analysis of multiple choice questions given to first year medical students-concept building and MCQs". *International Journal of Research in Humanities and Social Sciences*, 3(9): 18 – 29.
5. Chhaya, J., H. Bhabhor, J. Devalia, U. Machhar, and A. Kavishvar, 2018. "A study of quality check on multiple choice questions (MCQs) using item analysis for differentiating good and poor performing students". *Healthline Journal*, 9(1): 24 – 29.
6. Frantz, J.M., M. Rowe, D.A. Hess, A.J. Rhoda, B.L. Sauls, and L. Wegner, 2013. "Student and staff perceptions and experiences of the introduction of objective structured practical examinations: A pilot study". *African Journal of Health Professions Education* 5(2):72-74.
7. Gajjar, S., R. Sharma, P. Kumar, and M. Rana, 2014. "Item and test analysis to identify quality multiple choice questions (MCQs) from an assessment of medical students of Ahmedabad, Gujarat". *Indian journal of community medicine: official publication of Indian Association of Preventive & Social Medicine*, 39(1): 17-20.
8. Harden, R.M., and RG Cairncross, 1980. "Assessment of practical skills: The objective structured practical examination (OSPE)". *Studies in Higher Education* 5(2): 187-196.
9. Hingorjo, M.R., and F. Jaleel 2012. "Analysis of one-best MCQs: the difficulty index, discrimination index and distractor efficiency". *Journal of the Pakistan Medical Association*, 62(2): 142 - 147.
10. Izah, S.C., T.C. Odubo, V.E. Ajumobi, and K.E. Torru, 2021. "Item analysis of Multiple Choice Questions (MCQs) from a formative assessment of first year microbiology major students". *Research and Review Insights*, 5:1-6. doi: 10.15761/RRI.1000166.
11. Jaswal, S., J. Chattwal, J. Kaur, S. Gupta, and T. Singh, 2015. "Assessment for learning with Objectively Structured Practical Examination in Biochemistry". *International Journal of Applied and Basic Medical Research* 5:S71-S75.
12. Mahjabeen, W., S. Alam, U. Hassan, T. Zafar, R. Butt, S. Konain, and M. Rizvi, 2017. "Difficulty index, discrimination index and distractor efficiency in multiple choice questions". *Annals of Pakistan Institute of Medical Sciences*, 3(4):310 – 315.
13. Mard, S.A., and S. Ghafouri, 2020. "Objective Structured Practical Examination in Experimental Physiology Increased Satisfaction of Medical Students". *Advances in Medical Education and Practice* 11: 651—659
14. Mokkapat, A., G. Pavani, S.M. Dass, and M.S. Rao, 2016. "Objective structured practical examination as a formative assessment tool for IIInd MBBS microbiology students". *International Journal of Medical Science and Public Health*, 4(10): 4535-4540.
15. Munjal, K., P.K. Bandi, A. Varma, and S. Nandedkar, 2012. "Assessment of medical students by OSPE method in pathology". *Internet Journal of Medical Update*, 7(1):2-7.
16. Nigam, R., and P. Mahawar, 2011. "Critical analysis of performance of MBBS students using OSPE & TDPE—A comparative study". *National Journal of Community Medicine*, 2(3), 322-24.
17. Patel, R.M. 2017. "Use of item analysis to improve quality of multiple choice questions in II MBBS". *Journal of Education Technology in Health Sciences*, 4(1): 22 – 29.

18. Rao, C., H.K. Prasad, K. Sajitha, H. Permi, and J. Shetty, 2016. "Item Analysis Of Multiple Choice Questions: Assessing An Assessment Tool In Medical Students". *International Journal of Educational And Psychological Researches*, 2(4): 201.
19. Relwani, N.R., R.A. Wadke, S. Anjenaya, and P.N. Sawardekar, 2016. Effectiveness of objective structured practical examination as a formative assessment tool as compared to traditional method for MBBS students. *International Journal of Community Medicine and Public Health*, 3(12):3526-3532.
20. Vijaya, D.S., Alan, 2014. "Comparative Study to Evaluate Practical Skills in Physiology Among 1st Phase Medical Under Graduates At Jnmc Belgaum: Traditional Practical Examinations Versus Objective Structure Practical Examinations (TPE V/S OSPE)". *International Journal of Educational Research and Technology*, 5(1):126-134.
21. Vishwakarma, K., M. Sharma, P.S. Matreja, and V.P. Giri, 2016. "Introducing objective structured practical examination as a method of learning and evaluation for undergraduate pharmacology". *Indian Journal of Pharmacology* 48(Suppl 1): S47–S51.

Cite this Article: Izah, SC; Odubo, TC; Ajumobi, VE; Osinowo, O (2022). Item Analysis of Objective Structured Practical Examination (OSPE) Used as an Assessment Tool for First-Year Microbiology Students. *Greener Journal of Biological Sciences*, 12(1): 11-22.