How We Teach: Classroom And Laboratory Research Projects

MBBS student perceptions about physiology subject teaching and objective structured practical examination based formative assessment for improving competencies

K. Lakshmipathy
Department of Physiology, Employees State Insurance Corporation Medical College and Post Graduate Institute of Medical Sciences and Research, Rajajinagar, Bangalore, Karnataka, India

Submitted 30 June 2014; accepted in final form 12 May 2015

Lakshmipathy K. MBBS student perceptions about physiology subject teaching and objective structured practical examination based formative assessment for improving competencies. Adv Physiol Educ 39: 198–204, 2015; doi:10.1152/advan.00073.2014.—The objectives of the present study were to 1) assess student attitudes to physiology, 2) evaluate student opinions about the influence of an objective structured practical examination (OSPE) on competence, and 3) assess the validity and reliability of an indigenously designed feedback questionnaire. A structured questionnaire containing 16 items statements, 8 items on an Osgood’s 5-point semantic differential scale and 8 items on a Likert’s 5-point scale, was used. Options were assigned scores of 1–5 according to weightage. For Osgood’s semantic differential scale items, a χ²-test was done to analyze student attitudes toward the subject. For Likert scale items, mean score and SD were calculated to analyze student opinions of the OSPE. Item validity was assessed by item analysis, and reliability was assessed by calculating Cronbach’s α. The subject as a whole was interesting to 82% of the students (n = 135). The theory was interesting to 75% of the students (n = 132) but complex to 42% (n = 118). The practical was interesting to 93% of the students (n = 134); 76% of the students (n = 104) felt that the practical was simple, whereas 4% felt it was complex. The OSPE was interesting to 79% of the students (n = 131); 57% of the students (n = 116) felt it was simple, whereas 24% found it complex. Components of the subject, intricateness, and student interests were strongly associated. Students chose options on a higher weight scale, favoring the OSPE. Items were found to be valid and reliable. In conclusion, the subject of physiology was interesting but not simple to understand. Student interests varied with the components of the subject, and the components of the subject had varied intricateness. Students were in favor of the OSPE for assessment. The questionnaire used for the study was valid and reliable.

objective structured practical examination; assessment and learning; questionnaire: Osgood's semantic differential scale and Likert scale; item analysis: validity; cronbach's α: reliability.

TEACHERS must make scrupulous decisions when choosing student assessments. The most important parts of assessment are the interpretation and use of the information that is gleaned for its intended purpose. Assessment is embedded in the learning process (26). It is tightly interconnected with curriculum and instruction. The three attributes of assessment are assessment for learning, assessment as learning, and assessment of learning. Assessment plays a major role in how students learn, their motivation to learn, and how teachers teach and has powerful effects on student performance (12). Students become more focused and motivated if they are assessed for their capability of integration, application, and synthesis of knowledge and if their skills are observed and graded (11, 20, 24). The objective structured clinical examination is an assessment tool that assesses what students can do in a structured pattern objectively under direct observation and is able to assess the above-said capabilities (15). The objective structured practical examination (OSPE) is a modified form of the objective structured clinical examination but is used for evaluation of preclinical and paraclinical subjects (14).

A single examination at the end of an academic year does not fully assess a student. Conventional practical examination is relatively subjective and unstructured (6). The shortcomings of the different assessment methods have been well documented (7). The concept of OSPE is that there is no patient/test subject and examiner variability, so that the validity of the examination is improved (4). For complete assessment of competencies in clinical disciplines, Harden and Gleeson (14) recommended that the objective structured clinical examination should be combined with the traditional long case. Keeping this in view, the OSPE was introduced and standardized with a conventional practical exam at the All India Institute of Medical Sciences (30). The OSPE has yet to be implemented in many South Indian medical colleges and universities. Any change usually meets resistance (21).

Ensuring adequate feedback for students should be an important concern to review and revise the teaching programs (10). Teachers who are skilled observers, alert and able to know when a student needs correction, would not simply function as repositories of information or judges. Receiving the feedback requires maturity, honesty, and commitment to the goal of assessment. Validity and reliability are two fundamental elements in the evaluation of a measurement instrument (37). Validity is the extent to which an instrument measures what it is intended to measure. Reliability is the ability of an instrument to measure consistently on repeated measurements.

University rules stipulate a minimum of three internal assessment exams to be conducted in an academic year. Internal assessment exams are the routine conventional/traditional theory and practical exams, which are similar to university exams. It was observed during regular practical classes and internal assessment exams that students did not bother much on exact procedures while performing a skill. They were vague in answering questions, lacking focus and objectivity, thereby losing marks in internal assessment exams. Hence, the thought of introducing OSPE as a method of assessment came up, which would help students to be alert, very objective in
answering, and focused while performing skills. In addition, the OSPE aids in identifying areas of weakness of students so that corrective steps can be initiated and suitable teaching methodology can be planned without acute time constraints.

In the light of the above background, three OSPE sessions were conducted in the Department of Physiology. The feedback of students regarding the OSPE was taken for evaluation. The information obtained was to be considered in the decision of using the OSPE routinely for formative assessment in the absence of mandatory requirements by the affiliated university and to strongly advocate for its inclusion for summative assessment in university examinations.

The objectives of the present study were to 1) assess general attitudes of the students to physiology as a liking toward the subject influences the learning process to a great extent, 2) evaluate the perception of students toward the impact of OSPE in enhancing their competence, and 3) assess the validity and reliability of the indigenously designed feedback questionnaire.

METHODS

The present study was undertaken at a private medical college in Tamil Nadu, India. The medical program in India mandates 1 yr of study in physiology as a basic medical science subject along with two other subjects, anatomy and biochemistry, in the first year. The total program runs for 5.5 yr, leading to the Bachelor of Medicine and Bachelor of Surgery (MBBS) degree. One hundred fifty students who enrolled for the MBBS course at the institution in the academic year of 2011–2012 formed the study group. The present study was conducted in the Department of Physiology from October to December 2011. During this period, students faced three OSPE sessions and a conventional exam as part of the first internal assessment.

Ethics Committee approval was not needed as the study pertained to research on educational practices and was exempted by the Indian Council of Medical Research (17).

OSPE Sessions

For each OSPE session, the whole batch of students was divided into 2 batches of 75 students each. Each batch had the OSPE on 2 separate days. For each batch, there were 3 subbatches of 25 students. There was no mixing of students waiting for the OSPE with those that finished earlier.

There were two skill stations and six response stations in each session, with a total of eight stations. Three similar sets of these 8 stations were arranged in a big OSPE hall with a rest station (only when needed) so that 25 students could be accommodated at a stretch with 6 staff members allotted for 6 skill stations. Skill stations would repeat for different batches but not response stations. The portions included were those covered in the theory and practical classes. Skill station questions were regarding focusing of a red blood cell square, preparation of a peripheral blood smear, focusing a white blood cell counting square under low power, location of an apex beat, determination of blood pressure, etc. Response station questions pertained to theory aspects of topics like general physiology, muscular nerve physiology, blood, the respiratory system, and the cardiovascular system. Marks were compiled to a total of 40 marks, with 10 marks for skill stations and 30 marks for response stations.

In departmental meetings, elaborate arrangements required to be made for the conduct of the OSPE, framing of the questions, time allotment, and preparation of checklists/keys were thoroughly discussed to reach a consensus without any ambiguity. Marks were tabulated in an OSPE score sheet/grid sheet.

Before the first OSPE session, students were thoroughly oriented and sensitized regarding various aspects of the OSPE: questions, checklists, time allotted for stations, scorings, logistics involved, etc.

At the end of the third OSPE session, feedback was taken from all students to know their perceptions on the subject in general and the OSPE in particular using an indigenously designed structured questionnaire. This was subjected for statistical analysis to assess the credibility of the OSPE evaluation program.

Feedback Questionnaire

The feedback questionnaire was a structured questionnaire having 16 item statements. Each item had five options, which were given scores ranging from 1 to 5 according to weightage.

Prior instructions and adequate time to answer the questionnaire were given. Students’ anonymity was maintained by asking them not to write their names and roll numbers. One hundred forty students gave the feedback as 10 students remained absent.

Of the 16 items, 8 items belonged to Osgood’s 5-point semantic differential scale (OSDS), which measures the attitude of the students toward physiology as a subject. The remaining eight items were based on Likert’s five-point scale to assess the impact of the OSPE on various aspects of learning. They were analyzed for validity and reliability, which are the two important attributes of any measurement instrument. A valid and reliable questionnaire reflecting the opinions of students about the OSPE could scientifically prove the impact of the OSPE on their learning.

OSDS Items

OSDS items analyzed four important perceptions of the students to physiology as a subject: interesting/boring/simple/complex, which have direct bearing on their learning. Physiology as a subject was perceived 1) as a whole and also apportioned as 2) theory, 3) practicals, and 4) the OSPE.

The bipolar adjectives were as follows: very simple/very interesting (option A, 5 scores); somewhat simple/somewhat interesting (option B, 4 scores); neutral (option C, 3 scores); somewhat complex/somewhat boring (option D, 2 scores); or very complex/very boring (option E, 1 score), with the center being the neutral option.

Likert Scale Items

Likert scale items analyzed the impact of the OSPE on various aspects of student learning: 1) promoting sharp thinking, 2) focusing, 3) time spent to prepare for the OSPE, 4) uniformity in evaluation (without favorism/subjectivity), 5) preparing them to face the conventional exam with confidence, 6) having both conventional practicals and the OSPE, 7) opinion about canceling the OSPE, and, finally, 8) opinions about the method of conducting in the department. Items 1, 2, and 4–6 had the following options: strongly agree (option A, 5 scores), agree (option B, 4 scores), neutral (option C, 3 scores), disagree (option D, 2 scores), and strongly disagree (option E, 1 score). Items 3 and 7 had the same options but in reverse scoring, with the strongly disagree option getting five scores and the strongly agree option getting one score. Item 8 had the following options: excellent (option A, 5 scores), very good (option B, 4 scores), good (option C, 3 scores), fair (option D, 2 scores) and poor (option E, 1 score). A weighted score of five was given for options of having a very strong impact of the OSPE on learning, a score of one for options of having absolutely no impact on learning, and a score of three being neutral.

Statistical Analysis

SPSS software (version 20) was used for statistical analysis.

1 The feedback questionnaire can be found in the Supplemental Material for this article available at the Advances in Physiology Education website.
Scores of 16 items of the feedback questionnaire for every student were totaled. The total score of any student depended on the options chosen by them, which were given weighted scores. High scorers were those students who selected options of five/four scores for most of the items, and low scorers were those students who selected options with two/three scores for most of the items. High scorers felt that physiology was interesting and simple and that the OSPE had a positive impact in improving competencies. Low scorers felt that physiology was boring and complex and that the OSPE had no impact on learning.

Student scores were arranged in decreasing order of total scores; students in the upper portions showed higher scores than those in lower portions. The entire 140 student group was divided into 3 groups, with the upper third consisting of 47 students (high scorers), the middle third consisting of 46 students, and lower third consisting of the last 47 students (low scorers).

For each of the items, the exact number of students who selected options A–E were tabulated for all three groups. If any student omitted the options, it was noted. In this way, 16 tables for 16 items were obtained. By this arrangement, one can readily appreciate and compare the options chosen for an item by high scorers with low scorers. The computation of scores in this manner was needed for further statistical analysis.

Analysis of OSDS Items by Frequency Analysis and $\chi^2$-Test

The direction and magnitude of student attitudes to physiology was determined by the analysis.

Four items of the OSDS related to the interest of the students in the subject of physiology as a whole and its components. Students were subdivided into three groups: those who were interested, bored, or neutral. The frequency table obtained had four rows for the subject and its three components and three columns for the above-stated three groups. Their degree of association was analyzed by a $\chi^2$-test. In the present study, $P$ values of <0.05 were considered as significant.

The remaining four items of the OSDS pertained to the perceptions of the students to intricateness of the subject. Students were subdivided into three groups: those who felt it was simple, complex, or neutral. The frequency analysis table obtained had four rows for the subject and its three components and three columns for the above-stated three groups. Their degree of association was determined by a $\chi^2$-test. $P$ values of <0.05 were considered as significant.

Analysis of Likert Scale items by Mean Score, Item Analysis, and Cronbach’s $\alpha$

Mean score and SD. For an individual item, scores secured by all 140 students were totaled, and the mean score with SD was calculated. In a similar manner, mean scores and SDs were calculated for the remaining seven items. This showed the preferential scale of option for an item preferred by the whole class.

Item analysis. Item analysis was performed as previously described (5, 25). Item analysis is the process of analysis of an item after it has appeared in the exam/feedback. For each item, two indexes were calculated, namely, 1) difficulty index (also called facility value) and 2) discrimination index. If the items satisfy these indexes, the analysis can be subsequently repeated, and credibility can be assigned for the present exam/feedback of interest.

Item analysis in a conventional sense is applied to multiple-choice questions (MCQs) where one of the options is a correct answer. The difficulty index is the percentage of total students who answered a MCQ correctly. Item discrimination compares the number of students who answered a MCQ correctly in a high achiever group and a low achiever group. The higher the discrimination index, the better the item, because high values indicate a higher number of students choosing correct answers from the high achiever group. A negative value means that high achievers answered wrong and low achievers answered correctly, indicating a probable flawed item or a wrong key.

In a feedback questionnaire, there are no correct and wrong answers as such, and the option having the highest weighted score of five is the most preferred option, indicating the highest impact of the OSPE on learning. As the remaining options will not be a distractor in the real sense, the options not chosen by <5% of the students in the least will be labeled as nonfunctional options. The difficulty index of an item is the percentage of students (of the two groups) who preferred options with a score of five, meaning that they felt strongly about the impact of the OSPE on learning. The discrimination index of an item is the difference of the percentage of students who selected the option having a score of five between the high scorer group and the low scorer group.

**DIFFICULTY INDEX**. The difficulty index was calculated as follows:

$$ \text{Difficulty index (P)} = \frac{H - L}{N} \times 100 $$

where $H$ is the number of students who selected an option with a score of five in the high scorer group, $L$ is the number of students who selected an option with a score of five in the low scorer group, and $N$ is the total number of students who answered the questionnaire in the two groups. $P$ values were determined as follows:

$P \leq 30\%$ (low value) difficult items

($<30\%$ of students preferred an option with a score of 5)

$P > 30\%$ to $<70\%$ (medium value) moderately easy/difficult

($30–70\%$ of students preferred an option with a score of 5)

$P \geq 70\%$ (high value) easy items

($>70\%$ students preferred an option with a score of 5)

**DISCRIMINATION INDEX**. The discrimination index was calculated as follows:

$$ \text{Discrimination index} = \frac{H + L}{N} \times 2 $$

The range of the value is between $-1$ to $+1$. Accordingly, if the discrimination index is $<0.25$, items are okay; $0.25–0.35$, items are good; $>0.35$, items are very good; and $-ve$, items are bad.

An item discrimination index of 0.4 means that the number of students preferring the particular option is 40% more in the high scorer group than in the low scorer group or 40% less in the low scorer group than in the high scorer group. The item is able to discriminate between students in the high and low scorer groups to an extent of 0.4 (40%). In the present context of the feedback questionnaire, it implies that 40% more students in the high scorer group felt a positive impact of the OSPE on learning than in the low scorer group.

<table>
<thead>
<tr>
<th>Option A</th>
<th>Option B</th>
<th>Option C</th>
<th>Option D</th>
<th>Option E</th>
<th>Omit</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of students in the upper one-third</td>
<td>17</td>
<td>26</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>47</td>
</tr>
<tr>
<td>Number of students in the middle one-third</td>
<td>3</td>
<td>31</td>
<td>9</td>
<td>1</td>
<td>2</td>
<td>46</td>
</tr>
<tr>
<td>Number of students in the lower one-third</td>
<td>3</td>
<td>21</td>
<td>16</td>
<td>4</td>
<td>3</td>
<td>47</td>
</tr>
<tr>
<td>Total</td>
<td>23</td>
<td>78</td>
<td>29</td>
<td>5</td>
<td>5</td>
<td>140</td>
</tr>
</tbody>
</table>

*Option A*, strongly agree; *option B*, agree; *option C*, neutral; *option D*, disagree; *option E*, strongly disagree.
Consideration of an item for revision requires both indexes to be interpreted compositely. An item has to be considered for revision in two situations: 1) when the discrimination Index is −ve irrespective of the difficulty index value and 2) when the discrimination Index is between 0 and 0.25 and the difficulty index value is 30–70%.

Cronhbach’s α. Cronhbach’s α was determined as previously described (25, 36). Reliability, which is the best single measure of test accuracy, is the extent to which test results are consistent, stable, and free of error variance. It is the extent to which a test provides the same ranking of students when it is readministered, and it is measured by Cronbach’s α.

Factor analysis, which is the first step in the calculation of the α-coefficient, is a data reduction technique used to group various items having homogeneity/unidimensionality (36, 37). This reduced the eight items into three meaningful factorial groups, which were grouped as G1, G2, and G3 under the following headings:

- Perception of the impact of the OSPE on competency G1 Items 1, 2, and 5-8
- Time spent on the subject G2 Item 3
- Uniform evaluation G3 Item 4

For factorial group G1, Spearman’s rank correlation was calculated to check the interrelation of each item’s score with the total score (item interrelatedness) (36, 37). Cronbach’s α reliability coefficient was then calculated. Squaring α and subtracting from 1 gave the error variance (random error) (36).

The α-value normally ranges between 0 and 1. The closer Cronbach’s α-to 1.0, the greater the internal consistency of the items in the scale. Interpretation of the α-value varies according to the number of items. Reliabilities as low as 0.50 are satisfactory for short tests of 10–15 items, but tests with >50 items should have reliabilities of 0.80 or higher. If reliability is <0.8, a single test score should not be used to make important decisions about individuals (25). A high value of α (>0.9) may suggest redundancies (36).

Improper use of α can lead to situations in which either a test or scale is wrongly discarded or the test is criticized for not generating trustworthy results. The number of test items, item interrelatedness, and unidimensionality affect the value of α (25, 27, 35, 36). A low value of α could be due to a low number of items, poor interrelatedness between items, or heterogeneous constructs. If α is too high, it may suggest that some items are redundant as they are testing the same question but in a different guise. It has also been opined that without considering the quality of test items, increasing the number of items to improve reliability is not advisable (27).

RESULTS

Table 1 shows an example of an item (Likert scale item 1) arranged according to the options chosen by the students in three groups: the upper 1/3 (high scorers) of 47 students, middle 1/3 of 46 students, and lower 1/3 (low scorers) of 47 students.

Table 2 and 3 show the magnitude and direction of attitude of the students toward physiology Table 2 shows the distribution of students in terms of subject interest according to different components of the subject as felt by the students. Thirty-seven percent of the students (n = 131) were in favor of the OSPE as a learning method, whereas 8% were not in favor. The χ²-value for the data shown in Table 2 was 23.583, with a P value of 0.001, meaning that there was a significant association between student interest and the components of the subject.

Table 3 shows the distribution of students for the intricateness of the subject according to different components of the subject as felt by the students. Thirty-seven percent of the students (n = 118) felt that the theory was simple to understand, and 42% of the students found it to be complex. Seventy-six percent of the students (n = 104) felt that the OSPE was simple to learn, whereas 24% of the students found it to be complex. The χ²-value for the data shown in Table 3 was 55.645, with a P value of 0.000, indicating that there was a very high statistically significant association be-
tween different components of the subject of physiology and their intricateness.

Table 4 shows mean scores and SDs for each of the eight Likert scale items, the difficulty index, and the discrimination index along with nonfunctional options.

Mean scores of the items, with the exception of item 3, ranged from 3.25 to 4.11. Item 3 had a mean score of 2.9. Mean scores clearly indicated that most of the students chose options on a higher weight scale, showing an inclination toward the OSPE. Many options on the lower weight scale were nonfunctional, indicating that the OSPE had influenced their learning process.

Item 1, 3–5, and 8 had a difficulty index range from 6% to 25%, and items 2, 6, and 7 had a difficulty index range from 31% to 35%.

Item 2 and 5–7 had a discrimination index range from 0.38 to 0.46, meaning that they were very good items, item 1 was a good item with a discrimination index of 0.28, and items 3, 4, and 8 were okay with a discrimination index range from 0.12 to 0.22.

Table 5 shows a reference table of item analysis to determine revision of a Likert scale item, showing the overall assessment of eight items considering both the difficulty index and discrimination index in combination for each item (19). None of the eight items needed any revision. All items were valid.

Table 6 shows Spearman’s rank correlation coefficients with P values for each of the G1 items. P values were significant for all items at a 0.1% level of significance, meaning that there was an interitem correlation.

Reliability for Internal Consistency

Cronbach’s α calculated for G1 was 0.67 with an error variance of 0.55 \[1 - (0.67 \times 0.67) = 0.55\]. This was highly significant (25, 22).

Table 5. Reference table for item analysis

<table>
<thead>
<tr>
<th>Discrimination Index</th>
<th>Item Difficulty (P Value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;0 (bad)</td>
<td>Review</td>
</tr>
<tr>
<td>0–0.25 (okay)</td>
<td>Okay (items 3, 4, and 8)</td>
</tr>
<tr>
<td>&gt;0.25 (good)</td>
<td>Okay (items 1 and 5)</td>
</tr>
</tbody>
</table>

Table 6. Spearman’s rank correlation coefficient of each item with total score

<table>
<thead>
<tr>
<th>Item</th>
<th>Correlation Coefficient</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item 1</td>
<td>0.711*</td>
<td>0.000</td>
</tr>
<tr>
<td>Item 2</td>
<td>0.583*</td>
<td>0.000</td>
</tr>
<tr>
<td>Item 3</td>
<td>0.556*</td>
<td>0.000</td>
</tr>
<tr>
<td>Item 4</td>
<td>0.615*</td>
<td>0.000</td>
</tr>
<tr>
<td>Item 5</td>
<td>0.724*</td>
<td>0.000</td>
</tr>
<tr>
<td>Item 6</td>
<td>0.475*</td>
<td>0.000</td>
</tr>
</tbody>
</table>

*Correlation (significant at a 0.001 level of significance).

DISCUSSION

Assessment and learning are closely associated and interlinked. Assessment must take into account what is assessed, how it is assessed, and the assessment’s usefulness in fostering future learning (11). Within each domain of assessment, four levels have to be assessed: “knows,” “knows how,” “shows how,” and “does” (28). OSPE skill stations test shows how. Response stations in an OSPE test demonstrate knows (recall of facts, principles, and theories) and knows how (problem solving, application, and interpretation). Assessment for learning is where assessment helps teachers gain insights into what students understand to plan and guide instruction. Assessment as learning is where students develop an awareness of how they learn and use that awareness to adjust and advance their learning, taking increased responsibility for their learning.

Assessment of learning is where assessment informs students, teachers, and parents as well as the broader educational community of achievement.

Nayar (30) opined that the OSPE discriminates different levels of competence better than experiment examination. It has a high degree of reliability and validity. This was supported by a study by Mallick et al. (23). According to them, the OSPE tests a much wider sphere of subject matter as well as abilities of students than the conventional practical exam. It has scope for weighted representation of questions designed to assess different faculties (29). Our study is in complete agreement with this inference.

Mallick et al. (23) inferred that the majority of students showed a positive attitude to the OSPE and that high-rank students had a greater intensity of positive attitude. Students preferred a combination of OSPE and conventional practical examination. They became serious about attendance as well. In our study as well, students showed a positive attitude and improvement in attendance. A distinct advantage of introducing the OSPE was a change in students’ learning behavior. Their focus and concentration increased while demonstrating skills.

Aarti et al. (1), who conducted a study on the performance of students in different methods of examination of physiology at All India Institute of Medical Sciences, opined that OSPE marks were similar to clinical examination and different from graphs and charts. The OSPE can replace clinical examinations and not graph and charts. The OSPE can supplement but not replace conventional methods. Aarti et al. also felt that the OSPE may not be so useful in the final year course if investigation, differential diagnosis, and management are to be discussed.
A study conducted by Abraham et al. (2) at Mekala Manipal Medical College concluded that students were in favor of the OSPE compared with traditional practical examinations. A study at a Pakistan Medical College by Sandila et al. (33) found that the OSPE was an effective tool to discriminate between good and poor performers in physiology practical examinations. Another study at Lahore medical college by Hasan et al. (16) to assess the validity of the OSPE concluded that student attitudes toward the OSPE were found to be positive and appeared to be a valid index of learning attitudes of students throughout the year. The OSPE could fill in experimental deficiencies in renal, gastrointestinal, reproductive, endocrine, and electrophysiology. Gitanjali (13) formulated a semi-OSPE, an amalgam of conventional practical examinations and the OSPE, to overcome the burdens of the OSPE, such as time constraints, good human resources, observer fatigue, logistical problems, etc.

There is no gold standard for assessment (8). Clearly, no single test fulfills the criteria of a good examination, and the different methods complement each other (31). The criteria of a good examination include validity, reliability, objectivity, practicability, relevance, promotion of learning, power to discriminate between students, a relaxed environment, and positive student feedback. It was felt that the OSPE is one of the choices that fits into these criteria. Moreover, dismissal for incompetent performance in medical education is rare (9).

In the present study, we found that physiology as a subject was interesting to the students. Practicals were more interesting than theory. Students felt that theory was a bit complex. Students were very much interested in the OSPE. Regarding its simplicity and complexity, the results were a mixed bag. This attitude formed the basis of learning.

The χ²-value for the data shown Table 2 was 23.583, with a P value of 0.001, which means that there is a significant association between student interests and the subject. Stated more precisely, student interests vary with the various components of the subjects.

The χ²-value for the data shown Table 3 was 55.645, with a P value of 0.000, meaning there was a high statistically significant association between components of physiology and intricateness of the subject. Stated more precisely, the components of the subject affect its intricateness.

Mean scores for each Likert scale item were on a higher scale (Table 5). Most of the students felt that the OSPE makes them think sharply and helped them to focus on the subject and face the practical exam confidently. There was uniformity in evaluation, and students were satisfied in the way that the OSPE was conducted in the department. Some students felt that they were forced to spend more time with physiology and that their time would otherwise have been given to anatomy. Students felt that both conventional methods and the OSPE together would helpful in proper understanding of the subject. They were not for cancellation of the OSPE.

Each of the Likert scale item in the designed questionnaire of feedback was valid (Table 5). Validity is the most important consideration in test evaluation. The concept refers to the appropriateness, meaningfulness, and usefulness of the specific inferences made from test scores. Test validation is the process of accumulating evidence to support such inferences (3).

Cronbach’s α-value of 0.67 indicated that the questionnaire was highly reliable. Interitem relatedness, as shown in Table 6, was highly significant for all items. Reliability is the best single measure of test accuracy and means that the test results are consistent, stable, and free of error variance (25).

**Conclusions**

The students like physiology as a subject but feel that it is not simple to understand. Student interests vary with the components of the subject, and the components of the subject have varied intricateness. Students are in favor of the OSPE for formative assessment. The questionnaire used for the study was valid and reliable.

**DISCLOSURES**

No conflicts of interest, financial or otherwise, are declared by the author(s).

**AUTHOR CONTRIBUTIONS**

Author contributions: K.L. conception and design of research; K.L. performed experiments; K.L. analyzed data; K.L. interpreted results of experiments; K.L. prepared figures; K.L. drafted manuscript; K.L. edited and revised manuscript; K.L. approved final version of manuscript.

**REFERENCES**