

## **Development and Use of Checklists for Assessment of Medical Students in Clinical Examinations on Real Patients:**

### **A preliminary study**

Gamal E H El Shallaly<sup>1\*</sup>, Eltayeb A Ali<sup>1\*</sup>, Noha O Ahmad<sup>2\*</sup>

#### **Abstract**

##### **Background**

Clinical performance examinations (CPE) are important methods for assessing medical students and postgraduate medical trainees. Unfortunately, assessment (scoring) in these exams as it is practiced in most medical schools is based mainly on the general impression and level of experience of the examiner. Hence, it has the disadvantage of being non-structured, subjective, and sometimes highly biased. The objective of this study was to develop objective structured tools (checklists), use them in the clinical examination, and to compare their scores to those given by the conventional way.

##### **Methods**

Case-specific checklists were developed for some common long and short clinical examination cases. These were tried in the final undergraduate surgery exam at Alazhari University, August 2008. Scores (marks) given to the same student, using the conventional and the checklist systems were compared using the chi square and correlation statistics.

Questionnaires were filled by the examiners after using the checklists.

##### **Results**

Checklist and conventional scores were strongly correlated in the long case exams. A significant difference between the scores was, however, noted in the short case exams. Care should be taken in interpretation because of the small number of data. Evaluators felt that the checklist system was more objective and structured and hence fairer than the conventional method. Although checklist final scores took a bit longer time to calculate, that was not significant practically. Examiners gave valuable feedback regarding the construction and the use of checklists.

##### **Conclusion**

The use of checklists in the clinical examination was more objective, more structured, and more accurate than the conventional method. The development of checklists requires hard team work and frequent updating and use to develop experience. We propose using checklists as alternative tools of assessment with many advantages over the conventional method, and to prepare the examination culture to adopt the OSCE.

Keywords: performance, objective structured simulated patients.

**T**he clinical examination is an important and familiar form of assessment in Sudanese and other medical schools. It is also used in postgraduate medical or surgical examinations.

This format of clinical examination that involves real patients in a hospital set up is used by medical schools that have not adopted the objective structured clinical examination (OSCE).

Checklists are important components in the OSCE. Despite being more realistic than the OSCE, the clinical examination suffered from being non-structured, subjective, luck-dependent, and can sometimes be highly biased. Introducing checklists to this method of assessment has long been believed to be

---

1. Consultant surgeon, department of surgery, University of Alzaem Alazhari

2. House officer

\* Khartoum North Teaching Hospital

Correspondence to Mr. Gamal el Shallaly,  
email:gamalshallaly@hotmail.com

the way to rectify these disadvantages. The purpose of this study was to develop an objective structured tool-in checklist form-for use in the clinical examination and to compare it to the conventional way of scoring. To our knowledge, this is the first study of checklist use in the conventional clinical examination in Sudan.

### Methods

Clinical scoring checklists were developed for

a number of common clinical examination cases. These include long cases such as thyroid, breast, gall bladder disease, obstructive jaundice, and portal hypertension. It also included short cases such as inguinal hernias, ulcers, and lumps (Figures 1 and 2). The checklists were developed through a process of reviewing standard textbooks, consultation with experts, and previous examination experience.

Figure 1. Example of a Checklist History for Gall Bladder Disease (Long case)

### Date:

Name of Candidate:

Index No:

### History

#### Patient's information:

1. (name, age, sex, residence, occupation, marital status, parity, origin)

#### C/O- HPC

2. Abdominal Pain: Site (epigastric/Rt hypochondrial)
3. Duration
4. Radiation
5. Character: (colicky/stabbing)
6. Onset (sudden/gradual) & progress
7. Aggravating factors: (fatty food, spicy)
8. Relieving factors
9. Associated nausea or vomiting
10. Associated Fever & rigors: (Yes: cholecystitis/ No: biliary colic)
11. Associated Jaundice:
12. Progress of jaundice: (static, deepening, fluctuant)
13. Flatulence/belching
14. Dyspepsia/heart burn
15. Associated loss of appetite
16. Associated loss of weight
17. Change in bowel habits

#### Systemic enquiry:

18. Urinary system
19. Gynaecological system: Menstrual cycles (regular/irregular /amenorrhoea/ menopause)

#### PH

20. Similar condition (recurrent attacks)
21. Surgery (truncal vagotomy, resection of terminal ileum)
22. Haemolytic anaemia
23. Recurrent malaria
24. Blood transfusion
25. Diabetes mellitus & Hypertension

#### FH

26. similar problem (haemolytic anaemia/hepatitis contact)

#### SH

27. alcohol, smoking

#### Drug Hx:

28. Contraceptive pills

Figure2. Example of a checklist for lump examination (short case)

1. Greeting patient and introducing oneself
2. Asking permission from patient

**Inspection:**

Comments on:

3. Site (uses exact anatomical terms-distance from joints)
4. Shape
5. Size
6. Skin over it (color and texture)
7. Punctum
8. Veins
9. Pulsations

**Palpation**

10. Ask patient if it is painful before touching the lump
11. Comments on:
12. Tenderness: (starts with non-tender part/keeps an eye on patient's face)
13. Temperature: (feel with dorsum of fingers and compare to corresponding site):  
Comments: Normal/Hot/ Cold
14. Surface; smooth/irregular (boss elated, rough)
15. Edges: well/ill-defined
16. Composition:
17. Consistence: soft, firm, hard
18. Cross fluctuation (fluid)
19. Attachment to skin
20. Fixity to underlying structure
21. Pulsatile (transmits pulsation)/Expansile (aneurysms & very vascular tumours)
22. Compressibility (venous malformations)
23. Translucency (light passes easily across Clear Fluid)

**Percussion**

24. Fluid thrill

**Auscultation:**

25. Bruit (vascular)/ Bowel sounds (lumps/hernias containing bowel)

**Other important examination**

26. **Regional Lymph Nodes**

A questionnaire (Figure 3) was constructed for examiners to fill after using the checklists. A briefing on the purpose of the study and scoring (marking) system was given before the start of the exam. Examiners were invited to participate voluntarily. Those who agreed to participate were given a checklist booklet and a questionnaire, to be returned at the end of the exam.

The clinical examination format consisted of one long case and 3 short cases. A panel of

two examiners assessed a single student at a time. Examiners who assessed the student's performance in the long case were different from those who took him along the short cases. One examiner used the checklist, while the other followed the conventional way of assessment. The latter is based mainly on estimation and previous examiner's experience. For the next student, the examiner who used the conventional method earlier used the checklist and vice versa.

Name of Examiner (Optional) .....

1. Please fill the following table for each student you examine:
2. Compared to the non-checklist assessment, the checklist assessment is
  - a. Less objective
  - b. Of similar objectivity
  - c. More objective
3. Compared to the non-checklist assessment, the checklist assessment is
  - a. Less practicable
  - b. Of similar practicability
  - c. More practicable
4. Compared to the non-checklist assessment, the checklist assessment is
  - a. Less accurate
  - b. Of similar accuracy
  - c. More accurate
5. I found the marking system of El Azhari clinical checklist
  - a. Difficult to use
  - b. Easy to use
  - c. Very easy to use
6. Compared to the non-checklist assessment, the time taken to use checklists is
  - a. Longer
  - b. Similar
  - c. Shorter
7. I found the contents (10 subjects) of El Azhari checklist
  - a. Too little
  - b. Just enough
  - c. Too many
  - d. Recommends adding the following:

**General Comments:**

8. I think the strong points of El Azhari checklist are:
  - a. ....
  - b. ....
  - c. ....
9. I think its weak points are:
  - a. ....
  - b. ....
  - c. ....
10. I suggest the following to improve the checklist:

**Thank you**

Figure3. Questionnaire: Examiner's Evaluation of El Azhari Clinical Checklists

The checklist was divided into sections. These included: history, examination, investigation, and treatment. Marking of each section was done separately and a percentage calculated. The total mark of the case was the average of the sum of percentages of each section. A score below 50 is considered a Failure.

The checklist and conventional scores for each student were registered separately and independently. These were later compared using the Chi square test. Correlation statistics were done using Pearson's correlation coefficient.

At the end of the examination day, the examiners filled the questionnaires, and handed it to the research officer.

### Results

Ten out of twenty examiners agreed to participate in the study, and took checklists and questionnaires. Five out of those ten examiners returned the questionnaire; (25% response rate).

Data on scores were obtained on 13 students. Marks given to each student using the conventional and checklist system were tested using the Chi square test. There was an overall significant difference ( $p=0.038$ ,  $n=13$ ). Further analysis showed a significant difference in the short cases assessment ( $p=0.049$ ,  $n=4$ ), but not in the long case assessment ( $p=0.08$ ,  $n=9$ ) (table 1).

Pearson's correlation coefficient for all the 13 cases was 0.755, denoting a strong correlation at the 0.001 level (2-tailed).

Table1. Comparison of scores (marks) using conventional and checklist methods.

Student's No.	Case examined/ Long or Short e.g. Breast/long, Thyroid/short, etc	Marks using checklist	Marks using conventional non- checklist method
1	Submandibular swelling/ short	75	55
2	Inguinal hernia/short	60	65
3	Foot Ulcer (diabetic)/short	52	55
4	Collection of 3 short cases	38.6	35-40 (37.5)
5	Thyroid/long	87	75
6	Portal Hypertension/long	70	65
7	Thyroid/long	71	75
8	Portal Hypertension/long	80	75
9	Thyroid/long	40	48
10	Portal Hypertension/long	57	69
11	Thyroid/long	56	40
12	No data/long	64	65
13	No data/long	42	35

Total no. of cases= 13.       $df=12$        $X^2= 21.9$        $p=0.038$

No. of long cases =9       $df=8$        $X^2= 14.1$        $p=0.08$

No. of short cases=4       $df=3$        $X^2=7.85$        $p=0.049$

$X^2$  = Chi square

Pearson correlation=0.755. Correlation is significant at the 0.001 level (2-tailed)

Analysis of the examiners' evaluation forms showed that most examiners thought that the checklist system was more objective, and more accurate than the conventional system (table 2).

### Discussion

This paper tries to develop a solution to a problem long noted in the assessment methods used in the clinical examinations as used in many countries including ours.

Our final exam in surgery consists of a written exam, video-projected structured exam (ViPSCE), and a clinical exam. The written exam involves multiple choice questions (MCQs) and a problem solving structured exam. Both the written and ViPSCE assess all levels of the Knowledge domain, whereas Skills and Attitude are assessed by the clinical exam.

In our clinical exam format, the student is assessed using one long and three short cases. The student is allowed 35 minutes to work out the long case alone, unobserved. Two examiners, forming a panel, then discuss his/her findings in history and physical examination. These and other items including differential diagnosis, investigations and treatment options are discussed theoretically over 30 minutes. In contrast, the short case assessment is observed. Two examiners—different from those who examined the student on the long case—directly observe and assess the same student as he interacts with the patients. Thus the short case exam is the only exam assessing the student's clinical skills and attitude. The short case exam time is 20 minutes. The student examines an average of 3-4 patients.

Table2. Analysis of evaluation forms

Checklist compared to conventional is	Less objective (0)	Similar (1)	More objective (4)
	Less accurate (1)	Similar (0)	More accurate (4)
	Less practicable (2)	Similar (1)	More practicable (2)
	Difficult to use (2)	Easy to use (3)	
	Takes longer time (5)		
	Contents too little (3)	Just enough (2)	

In the end of both the long and short case exams, the student is given a mark (score) based on the global impression of the 2 examiners.

This conventional method of assessment is thus based on the examiner's level of experience and the standard of the first batch of the students examined. Clearly experience is extremely variable between examiners and so is the standard of the students. In addition there are certain psychological and personal elements that could lead to bias. There are always 'doves' that are too kind and generous in their scoring and 'hawks' who are extremely strict and not easily satisfied. Sometimes one of the 2 examiners is more senior and dominant than the other. For all

these reasons the clinical examination in this current format has been shown to lack validity and reliability and referred to as 'luck of the draw'<sup>1, 2</sup>. Attempts to improve and standardize the long case assessment include the objective structured long examination record (OSLER) developed by Gleeson<sup>3</sup>.

Checklists are instruments or tools of assessment that are objective, structured, reliable, and unbiased. They are used in the OSCE<sup>4</sup>, as well as in clinical skills training<sup>5</sup>. Different types of checklists have been developed<sup>2-5</sup>. Developing checklists is difficult. Doctors disagree on the contents and weighting. Furthermore, it has to follow an evidence based and be updated all the time<sup>6</sup>.

The clinical examinations in the format we use in most of Sudanese medical schools, is more like real life medical practice. The examination is held in hospitals on real patients (RP), in contrast to the standardized patients (SP) used in the OSCE. Although Waas et al didn't find a difference between the use of RP and SP in long case reliability<sup>7</sup>, assessment using RP is obviously more realistic and cheaper. We are also blessed by the fact that Sudanese patients are, in most cases, willing to cooperate and participate in the exams. The kindness of our patients should be encouraged by the system.

The advantage of the use of real patients in our clinical exam is marred by the lack of an objective tool of assessment. We, therefore, thought of developing case-specific checklists, and trying them in the clinical examination. The checklists include scores for knowledge, skills and attitude. The format is composed of history taking, physical examination, discussion and differential diagnosis, investigation and treatment options.

We did a small pilot study of the use of the checklists in our latest surgery exam of final year medical students, August 2008. The examiners were briefed on the study and the checklists and given a questionnaire.

The small response rate among examiners surgeons was disappointing. However, enough data were collected to give us a head start in this try-out phase.

There was a strong correlation between the scores of both checklist and conventional systems. This meant that both systems could work well, as long as the examiners were experienced and not biased. However, checklist system would appear to be fairer when a decision of fail/pass has to be taken. There was no significant difference between the checklist and conventional scores in the long cases. However, a significant difference was noted between the two scoring systems in the short cases. If there is a real difference, this may be due to the short time allotted to the student and/or the small number of short

cases, which may not be enough to give a good impression for conventional scoring. Nevertheless, care should be taken in interpreting this result because of the small number of the data used (alpha type error). More participation of surgeons and more data are required to get more robust statistics.

The examiners' evaluation forms analysis was encouraging. As most examiners thought the checklist system was more objective and more accurate than the conventional system. Its strong points were that it was standardized and fairer than the conventional way. The checklist is particularly more accurate in giving the deserved score than the global estimation used in the conventional system. This is particularly useful at both open ends of the grade. Thus, a student that examiners would feel he/she was excellent should give a mark of 75 or more in our grading open system, but how much? There is no tool in the conventional system to differentiate between the 75, 76, 80, or 90 mark, etc. The same occurs with failures below the 50 mark. Is it 49, 48, 40 or 30 etc? The mark here would also make a huge difference in the overall grade including the (Pass/Fail) status, since the total mark in surgery is the sum of adding all the different items of the exam.

The fact that it took about 2-3 minutes longer to do the final calculations was not really significant practically. Useful feedback included suggestions to increase the checklist contents by including more cases. It was also suggested that examiners should be trained before the exam on using the checklists.

Students should know that it is not enough to perform a certain skill but to do it correctly. Checklists are, therefore, not only useful tools in assessment, but could also be very useful ones in teaching and training.

In the future, we propose to change the current format of our clinical examination, especially dropping the unobserved single long case. We propose an observed multiple medium cases format, being assessed using our uniquely developed checklists.

While additional work on the development and use of checklists is clearly necessary, this study augments the beginning of a new checklist method of assessment of clinical performance examination in Sudan.

### Conclusion

There was a general strong correlation between the conventional and the checklist scoring systems. More data are however required. Overall, the checklist system seems to be more accurate in determining the actual scores, particularly at the open (excellent/fail) ends.

Compared to the conventional methods, the checklists system was thought by examiners to be more objective, more accurate, and easy to use. Checklist score calculation takes longer to do than the conventional method, but is fairer.

We recommend the use of checklist as a tool of assessment in the clinical examination to prepare the examination culture (examiners) to adopt the OSCE. It can also be useful as a tool for clinical skills teaching and training.

### References

1. Stokes J. The clinical examination - assessment of clinical skills, medical education booklet no. 2 (Dundee, Association for the Study of Medical Education.
2. Sood R. Long case examination. Can it be improved? *Journal, Indian Academy of Clinical Medicine* 2001; 2 (4): 251-55
3. Gleeson F. Assessment of clinical competence using the objective structured long examination record (OSLER). *Medical Teacher* 1997; 19: 7-14.
4. Harden RM, Stevenson M, Wilson DW et al. Assessment of clinical competencies using objective structured clinical examination. *Br J Med Educ* 1975; 1:447-451
5. M O'Comor, RC Mc Graw. Clinical skills training: developing objective assessment instruments. *Medical Education* 1997; 31 (5): 359-363.
6. Frank C. Evidence based checklists for objective structured clinical examination. *BMJ* 2006; 333: 546-548.
7. Wass V, Jones R, Van der vleuten. Standardized or real patients to test clinical competence? The long case revisited. *Medical Education* 2001; 35 (4): 321-25.