Perceptions of Teachers and Students on the Assessment Methods of Clinical Competence

Amha Mekasha*

Abstract: Focus group interviews were conducted to find out perception of students and teachers on the existing methods of assessing clinical competence at the Jimma Institute of Health Sciences. There was high concordance among the perceptions of students and staff regarding the characteristics of the various methods of clinical competence. In general, long case (LC) was a favoured method of assessment for its resemblance to the real clinical setting but was found to suffer from patient variability, subjective marking and lack of observation of skills by assessors. The short case (SC) was felt to be good because it is conducted under observation and sampling of cases is wider, however, it was criticised for its subjectivity and degeneration into theoretical discussion. Progressive assessment (PA) is favoured by most because it allows wider exposure to assessment and it is free from exam anxiety, however, needs to be objectified and used as a complement to the final examination. The viva voce (VV) was not found to be very useful by many as its conduct and purpose are not structured. Exams in general were felt to be essential in the assessment of medical competence, however, more emphasis should be given to the PA.

Introduction

In many parts of the world methods of assessment are developing along with the changes in curriculum of medical education. The functions expected of a graduating medical student are varied and complex. Consequently, it would be difficult to assume that a single method can assess these qualities of a competent practitioner. Katz and Snow (1980) said that a test is worthless however great its reliability or however objective it appears to be if it does not measure the qualities important in enabling the health worker to carry out his/her responsibilities. Hence, various methods of assessment are

* Associate Professor of Paediatrics and Child Health, Medical Faculty, Addis Ababa University.

being used among medical schools. The Jimma Institute of Health Sciences (JIHS) uses several methods of assessment including long case (LC), short cases (SC), viva voce (VV), progressive assessment (PA) in addition to MCQs. However, these methods have been under severe criticism for some time (Harden and Gleeson 1979, Gleeson 1996) elsewhere in the world.

Long case is a method of assessment whereby the student is given one real patient to take complete history and perform physical examination within 45 minutes. The aim is to measure the competence to identify and manage health problems using clinical information gathered. The candidate is examined by a pair of examiners. The advantages of this method are that it resembles real life situation and tests skill of arranging extracted data to make systematic presentations. However, the drawbacks are that the level of difficulty of the examination is not similar to all candidates, subjectivity in marking and inter-rater variability, and lack of direct observation during patient-physician interaction (Gleeson 1996; Amha 2004).

Short case is used to assess ability to elicit, recognize and interpret physical findings in a series of selected patients (Fleming et al , 1974). It allows the examiner's direct observation of the process, there is increased sampling of items and is conducted on real patients. The major criticism is that there may be shallow coverage of each case and suffers from inter-rater variability and subjectivity (Amha 2004).

Viva voce is a verbal interaction between the candidate and the panel of assessors and essentially assesses cognitive domain. Its main advantage is that the examiners can vary the question in wide areas. The main disadvantage is that it is affected by the personality behavior and verbal style of the candidate (Burchard et. al 1995)

PA is an assessment based on the observation of the student performance in real clinical settings by any combination of the staff

over the period of attachment. It is also called ward/clinic or in-training assessment. The daily observation of the "on-the-job" performance usually is given the greatest weight in student assessment (O'Donohue and Wergin, 1978). Its advantage is that it assesses skills which cannot be assessed in an examination setting; such as inter-personal and communication skills, attitudes and others. (Feleti et al 1974). It allows the students to obtain immediate feedback. However, PA cradles some limitations including subjectivity, assessment is also affected by extraneous variables such as the general behavior of the student.

Therefore, this study is initiated to explore the perceptions of students and staff on the various methods used to assess clinical competence at the JIHS.

Methodology

The study is cross-sectional in design using focus groups. The data were collected during the months of December 1996-January 1997.

The focus group was selected as the research instrument to collect information because the method is particularly useful for exploring people's knowledge and experience and can be used to examine not only what people think but how they think and why they think the way they do (Kitzinger, 1995).

Study Population

In this study focus groups of 4-6 participants per group were taken as suggested by Krueger (1994) because of the small number of staff and interns per department in the hospital. There was a total of 34 interns during the 1996/1997 academic year. They were divided into the four major clinical departments, namely Internal medicine, surgery, paediatrics and child health **(DPCH),** obstetrics and gynaecology **(OBGY)** and the Team Training Programmes (TTP) working at the health centres. Interns in the hospital were also

included in the study. Four staff members from each of the four clinical departments were involved in the interview. All the staff members were assistant professors except two who were lecturers. Therefore, four focus groups of interns and four focus groups of staff were formed each group representing the clinical department to which it is attached. Thus, eight focus groups were formed because two focus groups represented each of the four clinical departments where interns were attached.

We have selected interns because they are seniors among the undergraduates in that they have gone through all the courses and are well experienced with the different methods of clinical assessment. In addition interns have completed their clinical training and have done their qualifying clinical examination and will not have the fear to freely express their feelings compared to the lower clinical year students.

Data Collection

A one-hour briefing was given to the subjects about the purpose of the study and the methods of the interview. Clarifications were given during the discussion. Dates for the interview were then arranged according to the convenience of each group to achieve maximum attendance and participation.

The groups were interviewed for nearly an hour on a departmental basis on separate occasions by one of the investigators (AM). The time was set free of duty hours. The discussions were carried out in the offices of the Institute in a comfortable situation to avoid interruptions in the hospital and health centers. The interviews were recorded using a tape recorder.

Data Analysis

Recorded tapes of the group discussion were played to obtain a general impression and the flow of the discussion. During the

subsequent playbacks notes were taken by as many replays as required. The issues that arose were put onto a data sheet along with the questions. Then the summaries of the content were put into an overview grid as suggested by Knodel (1993). This was then analysed to condense similar issues into themes. Important points were taken as quotations with corrections for the English language structures.

Ethical Issues

Ethical clearnace was obtained from the research and publications committee of JIHS. During briefing the confidentiality of individuals and records was assured. It has been confirmed to the participants that the tapes will be destroyed after the data has been analysed. In all the discussions, every effort was made to avoid calling participants by their names during the recording.

Results

Perceptions of Students

Long case

Most of the students felt that common cases are selected for the LC but there were variations in the level of difficulty of the cases. All agree that cases are allotted to the candidate at random. In most departments the cases are fair but in OBGY the varieties are limited to, except the first two cases, ante-natal cases and this determines students' preparation.

Most students do not have written guidelines particular to the requirements concerning the exams. They deduce the important requirements from the way they were taught during bed-side teaching. They think the common features required by the examiners are to elicit relevant history, perform a physical examination, reach a diagnosis and suggest a plan of management. However, the inherent problem among examiners is the variation in their preferences. Some

may want to stress on management, others on history or on certain diseases, confidence or attitude. A student said, out that “We write the history according to the likes of the examiner if we know who is going to examine us.” In other words, students prepare differently for different examiners.

Most students said that the marking is subjectively influenced by the previous performance of the candidate, usually the PA. They said, “The marks awarded to us are beyond our expectations. We do not know what they expect us to answer. So we are worried only about passing the exam and not the scores.”

The majority of students believe that the LC is a good method to assess clinical skills for the forthcoming real life practice. However, there are certain worries about the LC in that students concentrate on the common cases and admitted patients only and ignore the uncommon cases. Other drawbacks were also mentioned as follows:

> There is variation in patients. Some students get simple cases while others may get difficult cases, though this is basically chance. The other problem is examiners’ varied view towards different students.
>
> When patients are not happy about the set-up of the exams they may turn out to be unco-operative to the candidate. Consequently, this may count against the student’s performance”.

These views can be summarised as variations in patients, level of case difficulty and examiner's views towards students.

Short case

Most students revealed that 3-4 cases of taught and common health problems are used. It was said, however, “In medicine, we cannot be taught everything, we have to know everything. Time or chance may not expose the student to all cases, but he/she has to know.”

There was a general view that the most important aspects of the short cases are performance of a physical examination and interpretation of findings. However, discussions often turn into deep theories rather than focus on the practice. SC also tests the speed of doing physical examination to reach a diagnosis covering wide areas. A student pointed out the advantages of the short case as follows: "The multiple cases give the candidate a wider opportunity in that if s/he one does poor in one case he can compensate in the subsequent cases."

However, there are some worries in that the majority said the limitations of SC are that the marking is subjective and observation by the examiners could worsen the anxiety of the candidate.

Viva voce

The general feeling of most students was that the questions for the viva voce are drawn from the taught areas and the selection is at random. Unanimously, they agree that the markings are subjective and depend on previous performances. Interestingly, most students do not bother about the viva voce because its weight is low in the final aggregate. Furthermore a student commented. "It is easy to tackle the viva because we can easily divert the question towards the area we are conversant with."

During the discussion it was noted that the strengths of the VV are that it assesses the level of confidence, communication skill and prompt thinking related to use of the knowledge base. It also allows the examiner to cover areas that have not been covered by the other methods of assessment. For example, in surgery the identification of surgical instruments is assessed in the viva voce.

Among the drawbacks commented by the students are that the language fluency and personality of the candidate could be factors that influence the marking.

Progressive Assessment

Most of the students know the required performances and they favoured progressive assessment because it shapes the development of the student. However, the question is its practicability. For instance, a student argued, “I may be in the ward whenever the assessor is not there or I may not be in the ward whenever the assessor is there, even though I spend most of my time in the ward. This situation may not be understood by the instructor and thus distort the effort of the student”. Another commented, “It is not progressive as the name claims. They (teachers) do not know the students.” On the other hand, it was perceived that the preceptorship used in the department of DPCH is beneficial because the instructors know the students assigned to them well.

Even though there are some practical constraints mentioned above most agreed that PA has merits in that it is carried over a period of time as opposed to a single encounter during examination which is dependent on chance. It also avoids the psychological stress caused by examinations. Furthermore, it assesses the student holistically. It was said“. It should be the most important way of assessment because in a one-day examination the student may be psychologically unstable to take up the case or may get a difficult case. It assesses the skill, knowledge and discipline throughout the attachment.”

Nevertheless, the PA is criticised for its subjective marking. All commented that non-academic and social relation issues may influence the marking. One of them said “If the instructor knows the student socially he may help the student; therefore, it needs strong ethical commitment.”

Finally, it was noted that there are differences among departments in their conduct of the PA. In some departments there is no continuous feedback to enhance improvement of the student performance. In another department the procedure is harassing, contrary to the

preceptorship used in DPCH where students are followed on their progress throughout the whole period of attachment.

General Perception about Examination

This study revealed that all students believed that there should be examinations because they encourage the student to work hard and the instructor to teach better. However, all felt that examinations are stressful. One of them said, “Death and examination will never be accustomed to and both are terrible. However, both need to exist.” It was also mentioned that, though there should be examinations, emphasis should be placed upon progressive assessment. A proponent about the PA said “The person who created examinations should be brought to justice because a one-day exam cannot guarantee the competence of the candidate.”

Perceptions of Teachers

Long Case

All examiners said that the selection of the cases is based on their relevance. The cases should be common health problems with revealing physical findings. Though in all the departments cases are allotted at random, one of the participants argued that “There is no significant difference among the cases selected. The similarity of the cases helps dissolve the degree of disparity that might occur with allocation of cases.”

In addition in all departments in order to maintain fairness a score is assigned to the student after discussion on their performance. The subsections of the performance are carefully analysed: the history, physical examination, analysis and investigation. In some departments the candidate is compared with the preceding candidate. Otherwise, there are no strict criteria of marking in any department. However, a staff commented, “The only criterion I know is to fail the candidate if found dangerous, i.e. not safe in interpretation and management of the case.”

In general the long case was perceived to be beneficial because it measures the clinical skills in a real life situation and tests the confidence of the candidate. On the other hand, instructors mentioned that the LC is subjective and suffers from patient variations. There could be problems due to exam anxiety and a language barrier in clerking the patients.

Furthermore, lack of checking on interpersonal and communication skill are appreciated. However, some instructors commented, “When there is a problem we ask the candidate to check on the history. It is checked by the disparity of the history given by the candidate and the history known to the examiner”. Others argued, “The communication skill should be assessed at other times than during the LC because there are language barriers and anxiety. It should be assessed during the PA.”

In addition, the LC, is limited to one case and this was mentioned as a concern. “It is a poor sampling method, giving one case out of a multitude of cases. It may not test the general ability of the candidate to handle varied clinical problems”.

In order to improve the LC, some instructors suggested increasing the number of long cases; and to improve the staff student relationship to eliminate their anxiety syndrome. To decrease personal bias, marking by multiple examiners using check-lists and use of more hospitals to increase the number and variety of cases have been suggested.

Short Cases

Most of the teachers stated that the SC is used to assess clinical skills and interpretation of findings using common health problems. One of the teachers mentioned, “We use relevant cases with findings that the student should pick on many body systems”. In general, in allotting cases the student will be exempted from a case similar to the LC. The variations of cases in allotment to candidates was not considered as a problem because, one of them mentioned, “There is

little variation of patients as all students go through the same or similar set of patients".

The marking system shares the same features of the LC. There are no marking criteria and marks are assigned based on the agreement reached after discussion among examiners on the performance of the student. One of the teachers said, "The general rule of thumb is whether the student has followed the procedures, picked the findings and interpreted correctly".

Interestingly, most examiners perceived SC as a useful method because it helps cover wider areas of skills and knowledge. One of the examiners said emphatically, "It helps to differentiate the real clinician from the bookworm." It has been noted also by some that it helps observe the approach of the student to the patient.

Nevertheless, the SC has been found to have some drawbacks as it is practised at the JIHS. It suffers from subjective markings and bias. The other problem is time shortage. It was said, " The short case is not natural; the student is not working at his own pace, he is under pressure by the presence of the examiners".

In general, SC has been perceived as a useful method of assessment. The drawback can be improved by training students through simulations of the exam process to abate the anxiety and using marking criteria.

Viva Voce

Different approaches of question selection among departments were found. For example, in internal medicine it is a lottery system where students pick up questions by lot. On the other hand in OBGY it depends on the case that was examined in the LC. If the case was obstetrics the oral questions would be gynaecological and vice versa. In DPCH the questions are usually based on the experience logbook which reflects common childhood illnesses and preventive paediatrics, neonatology and emergency paediatrics. In surgery the

questions are spontaneous and random. However, they include questions related to surgical instruments and the application of the instruments.

Though departments have varied approaches in question selection they have common approaches in marking whereby the marks are decided by discussion among the panel of examiners.

The VV has been in use because they felt that it helps cover content areas that were not touched in the other examinations and helps the examiner to go in depth to probe the knowledge of the candidate. One examiner described, “We can take the student to any corner of the subject, in contrast to other methods, including instruments.”

However, there are perceived limitations to the VV in that the marking is subjective and it tests only the knowledge base. For some students the barrier of communication skill affects their performance in spite of a good knowledge base. Consequently, one examiner asserted, “The iva voce is not a useful examination tool”.

Progressive Assessment

Most of the departments use marking criteria other than looking at the expected attributes. The marks are given individually by each assessor and the average is taken in OBGY. In Internal medicine and surgery the ward senior and the general practitioner assess the candidate by discussing the student’s performance. In DPCH the method of follow up and teaching is based on preceptorship. Consequently, two preceptors are responsible for the marking; in addition, it has to be agreed at a department meeting whether the score concords to the general staff impression about the student. In summary there are no consistent ways of marking. One of the instructors said, “We do not like the way we do it. We tend just to lump the students into groups of outstanding, very good, good etc.”

Nevertheless, it was pointed out, “The advantage of this system is

that it reduces bias because it is assessed on repeated occasions by different assessors." Many felt that PA is useful because it keeps the student in constant check and it is motivating to the student. Furthermore, timely corrective measures can be taken to prevent worsening of a problem. Also there is no examination anxiety that interferes with the performance of the student in the daily activities.

On the other hand, most participants mentioned that there was personal bias and lack of standards. For instance one examiner pointed out, "Smart ones show up when the teacher is around but the good ones work late at night in the ward when no supervisor is around . This method of assessment needs the devotion of the staff to avoid such mishaps".

General Issues Related to Exams

The preference for the best method of assessment is varied. However, most teachers felt that PA is good but another teacher opposed the PA because "A good proportion of the staff does not really know the students". Others felt that LC is the best as it is very basic to the practice of the practitioner if bias is excluded. Others preferred the short cases because "It covers wide areas of clinical skills and can also be observed." The least preferred was the VV because it is not standardised and examination time is too short.

In the final analysis, all the participants believed in the need of having examinations because "It remains as the only way of assessing the student. It also helps assess the teaching effects".

Discussion

Long Case

The cases selected for the LC and the objectives of testing the clinical skill on the whole patient in a real clinical setting offer the test to be highly valid. Similar to this study Gleeson (1996) asserts that in terms of theoretical construct validity, the LC is valid; however, in practice

due to lack of structure it frequently degenerates into viva examination. The major criticism about the LC is the inter-rater variability. The difference in rating may be because clinical teachers look for different attributes and have their own standards of measurements (Williams, et.al, 1996).

The assignment of scores by discussion prevents the problem. Newble (1991) concords with the JIHS practice that inter-rater reliability can be improved if marking is carried out by at least a pair of examiners deciding by consensus. Furthermore, the variation in the areas of preference and emphasis among examiners results in inconsistencies of marking. This drawback can be remedied by the use of check-list or rating scales (Newble, et.al, 1980). Nevertheless, Newble (1992) disagrees on this point by saying " once the check-lists become known to the students, they can be learned by rote with students being rewarded for thoroughness which may be unrelated to their ability to perform an appropriate clinical exam on real patients". In addition, Van der Vleuten et al (1991) said that such objectified methods do not inherently provide more reliable scores and may even provide unwanted outcomes, such as negative effects on student behaviour and triviality of the content being measured". Finally, the use of check-list in reality prevents the examiner from moving out of the question when further elaboration is needed during the discussion of the case.

The major concern about the LC is the case allocation, selection and the marking systems. There appears to be general guidelines in the selection of cases but the LC does not allow the examiners to adjust for the degree of difficulty posed by patient at interview (Price and Byrne 1994). These authors found significant correlation between the examiners' and students' perceptions on case difficulty ($r = 0.438$, $P < 0.0001$) indicating that examiners can have similar perceptions to students about case difficulty.

Over the past few decades standardised patients have been used in some parts of the world to remedy case variations (Van der Vleuten and Swanson, 1990). The preparation of a standardised patient incurs heavy economic demand. It needs about 30 hours to train a standardised patient (Cusimano et al, 1994) and the cost per patient in some studies was very high (Stilman et al, 1986). The possibility of training standardized patients is, at the moment, remote in Ethiopia because of economic constraint.

The LC is a very useful method in assessing clinical competence in particular if the process is observed. Observation allows the examiner to assess the attitudes and communication skills during candidate-patient interactions. The assessors can also appreciate the attitude and co-operation of the patient during history taking and physical examination (Gleeson, 1996). However, it is important to consider the psychological make up of the student population because the examination anxiety may worsen in some students by the presence of the examiners. Furthermore, observing a student through the whole process is beyond reality, at least in the Ethiopian context where the doctors are already stretched.

In summary, the LC can be improved at the JIHS by using carefully selected cases on the basis of significance of the health problem, examined by a pair of examiners using a brief and comprehensive check-list under observation.

Short Cases

The major criticism of the SC has been the subjective marking similar to the LC as mentioned above. However, the marking can be objectified as mentioned earlier in the LC. The psychological turmoil the students encounter during observation should carefully be managed so that students do not get more anxious. During the regular classes browbeating of the students should be avoided and encouragement and motivation should be upheld.

As mentioned above in SC there is a drift towards theoretical discussion and often there is no room for testing ability to perform procedures (Paul 1994). Increasing the practical tests by using more cases and performing some feasible procedures using mannequins can increase the validity and the reliability of the method.

Similar to previous studies shortage of time during skills tests should be seriously considered. Van Luijk et al (1990) reported in their study that 65% of the students in skill tests were short of time to demonstrate their skills. This could be due to students skills, patients compliance or because adequate time has not been considered well for the exam-strained student.

Viva Voce

Paul (1990) argued that the VV is useful to assess communication skills and confidence level, but in the study of Wakeford (1985) examiners felt that the VV is generally useless. Hence, because of its lack of consistency as to what attributes it is designed to measure, the place of VV in the assessment of clinical competence remains questionable.

Progressive Assessment

The emphasis laid upon and the objectives of PA varies among medical schools. Therefore, its contribution and relation to other forms of assessment varies. The PA is considered to have high validity because of interaction with actual patients in real time using a wide range of observations regularly on many essential areas of clinical competence which may be difficult or impossible to assess in examination settings (Felleti et al, 1994). The same concept prevails among our students and examiners but the major problem is the practicality of carrying out observations. For instance, Stillman et al. (1986), carried out research involving 14 internal medicine residency programmes in the USA and reported that almost 20% of the residents in internal medicine had never been observed and about

35% had been observed only once or twice. In fact, the problem becomes worse if there is a mismatch between the number of students and teachers.

Progressive assessment should be objectified to strengthen its reliability. The preceptorship method in **DPCH** does not need extra manpower. In addition, because of the enhanced intimacy with the instructor the psychological turmoil that may occur during the observed clinical tests will be minimised. Moreover, learning is most effective when threat is reduced to a minimum.
Continuous and immediate feedback serves to fix the learning to make it permanently available. Therefore, the feedback should be improved and conducted on a continuous basis with the aim for the student to improve upon his or her weaknesses and keep up strengths. In order to keep track of the progress of each student there should be a structured record agreed upon by both parties. This record would also be helpful to make the marking more objectified.

It should be emphasized that PA is not merely present in the ward; it includes progress of the students in his/her knowledge, clinical skills, team work, attitude, personality, case discussions, seminar presentations and side-lab activities.

General Issues about Examination

Examinations are formidable and omnibus approaches to certification, with advantages and disadvantages. In this study there is a general feeling among the students and the staff that exams are necessary, though not liked by the majority, because there is no other way of certifying doctors at the moment. Examinations can assess limited attributes that do not guarantee the quality of the candidate (Felleti et al 1994). Furthermore, the time, cost and intensive labour are seriously limiting in the context of the JIHS. Therefore, until a better way of assessing clinical competence is discovered, clinical examinations will remain as the mainstay of assessment.

Conclusion and Recommendations

The cry for change is not an illusion but a real need to discover the best methods of assessment endowed with high validity, reliability and practicality. However, with the limited knowledge available some changes can be proposed to alleviate the existing tension. Therefore, based on the data presented and the literature review undertaken the following are recommended to improve the methods of assessment of clinical competence.

First there should be examinations until a valid and reliable in-training assessment method that would enhance the learning-teaching process is implemented.

The PA should be considered as the mainstay of assessment if adequate staff commitment is attained, and if it is objectified and structured. It allows staff to assess the candidate over a wide range of clinical problems in real-life clinical settings, and allows to examine skills that cannot be assessed in examination settings.

The LC can be improved by observing the whole process instead of leaving the candidate alone. Time for patient examination can be reduced because most of the tasks are observed during the clerking time. The panel of examiners should use carefully designed check lists that include the relevant tasks required of the candidate. To increase the sampling of clinical cases the SC should continue with as many cases as possible. It is suggested that the VV should be removed as all the attributes it measures can be covered by the other methods and it also adds little to the assessment process. Objectified structured clinical examinations (OSCE) as suggested by Harden and Gleeson (1979) are being used by many medical schools as a solution to the drawbacks of the traditional methods. However, OSCE has the problem of compartmentalisation of skills, time consuming to set up, and problem of space and economic constraints. Though, it seems impractical at the moment that the existing methods, as mentioned above can be improved.

References

Amha M.(2004). Assessment Methods in Medical Education. Ethiop Med J, 42, 63-71.

Burchard KW, PA. Rowland-mourin NP. Coe and JL.. Garb (1995). A Survey of Oral Examination: Inter-rater Agreement and the Influence of Rater Characteristics. Academic Medicine, 70, 1044-6.

Cusimano, MD., R. Chen, and W. Tucker (1994). *A Comparative Analysis of the Cost of Administration of an OSCE*. **Academic Medicine**, 69, 571-576.

Feletti, G., D. Cameron, B. Dawson-Saunders, J. Des Croseilliers, B. Dooley, Farmer, E., and McAvoy, P. (1994). In-training Assessment. In Newble, D. et al (Eds.). *The Certification and Recertification of Doctors*. Cambridge: Cambridge University Press.

Fleming, PR., Manderson, WG., Mathew, MB., PH., Saneson, and JF. Stokes (1974). *Evolution of an Examination: MRCP* (UK). **British Medical Journal***, 2, 99-107*.

Gleeson, F. (1996). Long Case Assessment Using the Objective Structured Long Examination Record (OSLER). *Archivos De La Facultad De Medicina de Zaragoza*, 36, 31-33.

Harden, RM., and F. Gleeson, (1979). *Assessment of Medical Competence Using Objective Structured Clinical Examination*. **Medical Education,** 13, 39-54.

Katz, FM., and R. Snow, (1980). *Assessing Health Workers Performance*. **Public Health Paper** No. 72. Geneva: WHO.

Kitzinger, J. (1995). *Introducing Focus Groups*. **British Medical Journal,** 311, 299-302.

Knodel, J. (1993). *The Design and Analysis of Focus Group Studies. In Morgan*, DL (ed.), **Successful Focus Group***: Advancing the state of the art*. London: a Sage pub.

Krueger, RA. (1994). *Focus group: a Practical Guide for Applied Research*. London: Sage

Newble, DI., J., Hoare, and PF. Sheldrake, (1980). *The Selection and Training of Examiners for Clinical Examination*. **Medical Education,** 14, 345-349.

Newble, D. (1991). *The Observed Long Case in Clinical Assessment*. **Medical Education**, 25, 369-373.

Newble, DI. (1992). *Assessing Clinical Competence at the Undergraduate Level*. **Medical Education**, 26, 504-511.

O'Donohue, WJ., and JF. Wergin, (1978). *Evaluation of Medical Students During Clinical Clerkship in Internal Medicine*. **Journal of Medical Education***,* 53, 55-58.

Paul, V. (1994). *Assessment of Clinical Competence of Undergraduate Medical Students*. **Indian Journal of Paediatrics,** 61, 145- 151.

Price, J., and JA.**,** Byrne (1994). *The Direct Clinical Examination: an Alternative Method for the Assessment of Clinical Psychiatry Skills in Undergraduate Medical Students*. **Medical Education**, 28, 120-125.

Stillman, P., DB., Swanson, S, Smee, (1986). *Assessing Clinical Skills of Residents with Standardised Patients*. **Annals of Internal Medicine**, 105, 762-771.

Van der Vleuten, CP., and DB. Swanson, (1990). *Assessment of Skills with Standardised Patients: State of the Art*. **Teaching and Learning in Medicine**, 2, 58-76.

_____ GR., Norman, and ED., De Graaff, (1991). *Pitfalls in the Pursuit of Objectivity: Issues of Reliability*. **Medical Education**, 25, 110-118.

Van Luijk, SJ., CMP., Van der Vleuten, and RM., Van Schelena, (1990). *Observer and Student Opinion about Skills Tests*. In Bender, et al (Eds.), **Teaching and Assessing Clinical Competence**. Groningen: Boekwerk. Pub.

Wakeford, R. (1985). *Examiners' Beliefs about the Utility and the Appropriateness of Different Examinations Techniques*. **Medical Education,** 19, 505-506.

Williams, C., P., Trigwell, and D. Yeoman, (1996). *Pass the Royal College Examinations. MCQ Technique*. **British Journal of Hospital Medicine**, 55, 479-481.