

EDUCATION

Assessment of the Psychomotor Domain

Amina Hamed Alobaidi, Kirkuk University College of Veterinary Medicine, Kirkuk, Iraq [KUCOVM],

Email: aminahamed2006@gmail.com, Mobile: +9647702399175,

ORCID: <http://orcid.org/0000-0002-8495-3452>

Received: 16/2/2020 Accepted: 22/4/2019 Published: 1st May, 2020

Introduction

With the introduction of modularization, many of medical colleges have moved to a learning outcomes based approach to ensure that curriculum design evolves from a more teacher-centred (content) to a more student –centred (learning) focus. Identifying learning outcomes enables both the teacher and student to clearly identify what a student is expected to have achieved or have made progress towards achieving on completion of a module [1].

A range of methods for competence (performance) assessment exist. Many methods for the assessment of psychomotor domain were developed to assess the communication and clinical skills, procedural skills and professionalism. Advances in psychometric theory and technology have been paralleled by the development of assessment instruments that improve the evaluation of these skills [2].

Taxonomy of psychomotor domain

Learning taxonomies or classifications are commonly utilized as a way of describing different kinds of learning behaviours and characteristics that we wish our students to develop. Different learning stages development identified using learning taxonomies and a useful tool for distinguishing the appropriateness of particular learning outcomes for particular module levels within the programmes are provided by such classification. This domain is characterised by progressive levels of behaviours from observation to mastery of a physical skill. Several different taxonomies exist. The most common

and earliest of these is Bloom's Taxonomy [3], adapted more recently by Anderson et al [4][Fig.1]. However, there are another versions of psychomotor domain taxonomy.

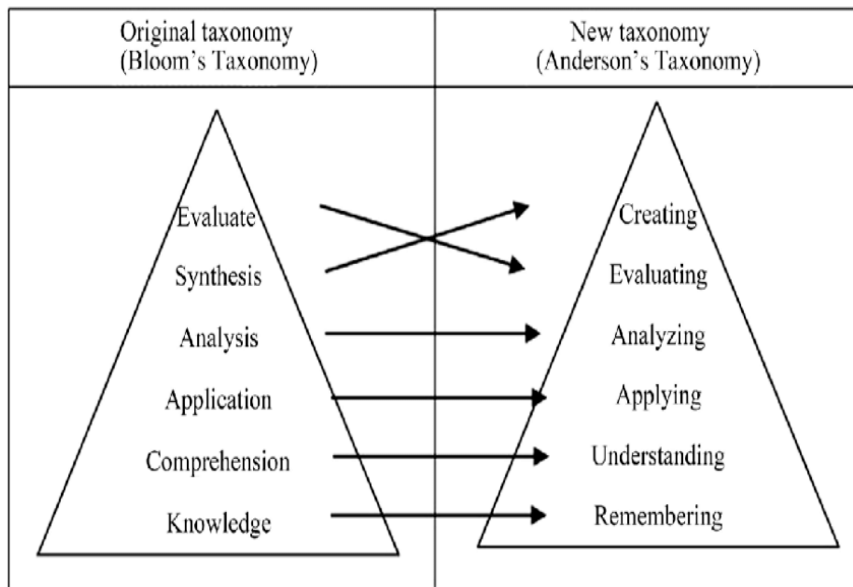


Fig. 1. Bloom Taxonomy of cognitive domain and revised form by Anderson [4].

Dave's [5] Taxonomy:

Table 1 shows the details of Dave's Taxonomy levels, categories, behaviours description and some verbs that used in objectives formulation. Dave's psychomotor domain as described in the table above is probably the most commonly referenced and used psychomotor domain interpretation

Simpson Taxonomy

Simpson [6] built this taxonomy on the work of Bloom and others. This version is particularly useful if we take adults out of their comfort zones, because it addresses sensory, perception, and preparation issues. Also it is appropriate for other types of adult development. In such cases Simpson's model would be more appropriate than Dave's.

Simpson's interpretation of the psychomotor domain differ from Dave's chiefly because it contains extra two levels prior to the initial imitation or copy stage. However, the first 2 levels of Simpson's 'Perception' and 'Set' stage are incorporated within the first 'Imitation' level of Dave's, assuming that you are dealing with fit and healthy people (probably adults rather than young children), and that 'getting ready' or 'preparing oneself' is part of the routine to be taught, learned or measured. If not, then the most comprehensive Simpson's version might help ensure that these two prerequisites for physical task development are checked and covered. As such, the Simpson model version is probably preferable than the Dave model for the development of young children.

Harrow's Taxonomy

Harrow's [7] interpretation of the Psychomotor domain is strongly biased towards the development of physical fitness, dexterity, agility, and control of the physical 'body', to a considerable level of expertise. As such the Harrow model is more appropriate to the development of young children's bodily movement, skills, and expressive movement than, say, the development of a corporate trainee's keyboard skills. By the same token Harrow model would be perhaps more useful for the development of adult public speaking or performance than Simpson's or Dave's, because the Harrow model focuses on the translation of physical and bodily activity into meaningful expression. Emotional influence on others is implied by only Harrow model of the three Psychomotor Domain models.

Bloom's taxonomy is a wonderful reference model for all involved in teaching, training, learning, coaching- in design, delivery, and evaluation of these development methods [4,8]. At its basic level the taxonomy provides a simple, quick, and easy checklist to start to plan any type of personal development. It helps to open up possibilities of all aspects of the subject or need concerned, and suggests a variety of the methods available for delivery of teaching and learning. As with any checklist, it also helps to reduce the risks of overlooking some vital aspects of the development required [1,9]. Each domain detailed elements in Bloom's taxonomy added reference information for evaluation and design of learning, whether for whole organization, large students or trainees group, syllabus or program,

entire course, single lesson, session, or activity, or training need. Taxonomy of Bloom's is continuously evolving, through the work of academics following in the footsteps of Bloom's early associates, as a fundamental concept for the development of formalised education across the world. Whatever the taxonomy, it seems to me that the list presented in Table 4. is a combination of the whole taxonomies. The behavioural verbs appropriate for the psychomotor domain are shown in Table 5.

Methods of Assessment of Psychomotor Domain

The use of a variety of different assessment methods has been characteristics of medical education , credentialing, and license since 1950s and then there has been a rapid and extensive change in the way assessment is conducted in the medical education based on a number of developments [2]. The selection of assessment method have rested primarily on validity and reliability [10,11]. Recently, van der Vleuten and Schuwirth [12] have expanded these factors for purpose of assessment in medical education and added acceptability, feasibility and educational effects.

The psychomotor domain includes coordination, physical movement and use of the motor skill areas. Practice is required for these skills development and their measure achieved in terms of precision, speed, distance, procedures or techniques in execution. A range of methods for psychomotor domain exist. However, in Iraq, most of the medical colleges run a traditional curricula and assess their students skills and knowledge by using written and oral examination. But TUCOM, is the only Iraqi College that adopt innovative curriculum and practice a different type of assessment. The effective student assessment program will include a variety of procedures [1]. The following are some of the more commonly used procedures with the indication of their main functions:

1. Informal observation:

This include any assessment tool that involves directly observing students as they utilize the skills taught in the class and the evaluation of student work that was produced by directly using the skills taught in the class [1]. The strength of this assessment form is that it provide direct measures of students skills and can involve real problems similar to which students exposed in the professional training of the

discipline. This method may have reliability and/or validity problems unless good rubrics are used since evaluation of student work of some types may have a strong subjective components [13]. The method may be used during instruction and alerts the too many specific guidance ways for student learning [1]. The ongoing observations and instructional decisions are a basic means of improving student learning [1].

The observation assessment method works well for assessment of psychomotor skills (use of laboratory equipments, population demonstration capabilities, etc.). However, it required careful implementation in order to make it easier to produce quantitative data for outcomes assessment reporting [14].

2. Classroom achievement test

This test should provide a systematic procedure for determining the extent to which intended learning outcomes have been achieved [1]. Teachers need to have a clear understanding of types of test items [tools], test assembly, test administration, testing practice, validity, reliability [15]. Item type must be carefully considered when tests are being constructed [13]. A major decision is the relative percentage to be assigned constructed response item (essay and completion) as opposed to selected response item (MCQ, T or F, matching). A variety of learning outcomes (factual knowledge, understanding, application, reasoning ability, problem solving) can be measured by this test method if it is performed properly [1]. However, this test was used primarily for measuring the recall of factual knowledge and have been widely criticized for the limited measure of achievement they provide. [1]. Some suggest to eliminate this testing from student assessment. However, I think to retain the test for assessment, but there is a need to change its construction which may lead to measurement of a broad band of achievement. The classroom test requires careful planning in order to provide direct measures or student skill or involve "Authentic" problem. Exam questions can be linked directly to outcomes for assessment purposes. Generally, the tests are quite acceptable to measure achievement of educational outcomes. However, it is recognized today tests are more useful at some educational settings than others. It is also recognized test results alone rarely provide a comprehensive picture of achievement. To provide a

complete picture, test results should be used in conjunction with evaluation data gathered through the use of other sources. Tests are practicable, but care must be taken to sample the domain correctly and to create rigorous items. Care must be taken to ensure reliability and validity if the test is to have value as an evaluation tool [13].

In a Malaysian higher learning institution the classroom assessment practices among lecturers was an effective method for the cognitive domain [16]. It provided evidence for teacher assessment and peer review assessment that may be used for to guide institutional leaders decisions regarding selection of proper classroom assessment methods [17,18,19].

3. Performance Assessment:

This method is needed when performance skills are not adequately assessed by paper –and–pencil tests alone [1]. In medical education, performance assessment provides a useful adjunct to paper-and-pencil tests. Ongoing performance skills observation and adjustment is usually obtained by using checklist, rating scale or holistic scoring rubric [1]. Performance assessment purpose is to direct the observation toward the most important elements of the performance and to provide information for judgments decision [1]. The strengths and limitation of performance assessment are shown in Table 6. The task of performance may be restricted to fit a specific and limited skill (e.g. measurement of IgE level in serum) or extended to a comprehensive performance that includes numerous specific skills (e.g. the predicting of biomarkers role in diagnosis of allergy). Effective performance assessments should be performed in systematic approach and follow the following outline steps [1].

- Specifying the performance outcomes.
- Selecting the focus of the assessment – Assessing the procedure and assessing the product.
- Selecting appropriate degree of realism.
- Selecting the performance situation :[Paper-and-pencil performance, Identification test, Structured performance test, Simulated performance, Work sample, and Extended research project]
- Selecting the method of observing, recording, and scoring.

The common procedures used are [1]: Systematic observation and anecdotal records; Checklists; Rating scale; and Scoring rubrics. Performance assessment in medical education provide a useful information concerning student achievement. However, they are subject to all of the observation errors and judgement errors (tendency to overrate) and halo effect. Thus, special care must be taken to improve the objectivity, reliability, and meaningfulness of the results. Table 7 shows the ways to improve the usefulness of performance assessment. Reported studies indicated the effectiveness of performance assessment in professional learning [20-28].

4. Product assessment

This is actually a form of performance assessment, but some types of performance result in a product that it forms the main core of the assessment [1]. In medical education the example for this are , Field work research project, oral presentation of a research work, Map design of a model or educational course, poster presentation, and laboratory report, and thus these products can be assessed directly. Checklists, rating scales, and holistic scoring rubrics are also useful in product assessment [1].

5. Portfolio assessment:

Portfolio is a collection of one's professional or personal goals, achievements and methods of achieving these goals [29]. Portfolio contents are ones best essays, written or research projects, logbook, letter of reflection, and evidence of professional growth [30]. The portfolio is with the following advantages [13]; Collect evidence of actual performance in the "Does" level in a longitudinal manner; Highly valued as a formative assessment and feedback tool. Although portfolio is an important assessment method [31], it is with limitations, such as time consuming on both faculty and students; Marking and standardization difficulty; Difficult to decide on pass/fail cut-off [13]. Portfolios are becoming an important means of assessment in medical education. It is used in some cases as a basic or sole, method of performance assessment. In others, they provide another useful tool in the teacher assessment kit [1]. The portfolio is either a developmental portfolio [organized to show student learning progress] or showcase portfolio [show samples of the student best

work [1]. Thus developmental portfolio is to be used throughout an instructional program and the showcase portfolio at the end of the program.

Portfolio assessment value is found in the vast array of evidence of learning they provide, the actual use of students sample of work, the active participation of students in selecting entries and maintaining the portfolio, and the variety of types of assessment data included [1]. In order to introduce effective planning of portfolio it is essential to determine the purpose, entries to be included, the guidelines for selecting and evaluation the entries, the procedures for maintaining and using the portfolio, and the criteria for an overall evaluation of the portfolio [1].

Portfolio provide a comprehensive assessment of student achievement and provides an opportunity for students to reflect on what is being learned. Furthermore, portfolio in addition to its main purpose [Improve student learning], it encourage student to participate more actively in the learning process and become more responsible for their own learning [1]. Thus in innovative curriculum the portfolio act as assessment method and a tool to improve skills. A series of questions concerning portfolio makeup, organization, and content may provide a structural evaluation of portfolio. Rating scale that focuses on learning outcomes being assessed by the portfolio may be used for overall evaluation of student progress (developmental portfolio)[1], while the final level of student performance (showcase portfolio) can be best determined by holistic scoring rubrics for each of the major areas of instruction included in the portfolio [32].

6. Simulation:

It may be considered as a performance assessment method. However, due to its widely using in medical education I considered it as an assessment method entity. Simulations are increasingly being used in medical education to ensure that examinees can demonstrate integration of prerequisite knowledge, skills and affect in realistic setting [33]. This reflects that subject matter experts concerns that traditional and selected response item formats measure knowledge, but not performance. In medical education, both standardized patient simulations [34,35] and computer simulations [36,37] are widely used to assess examinees clinical skills and medical problem solving.

Simulation have been implemented in various testing programs based on belief that fidelity is as important a feature of assessment as reliability and that the realistic challenges they pose increase their acceptability to examinees and content experts and enhance examination validity. Research has shown that it is feasible to develop assessments that produce scores that meet reasonable psychometric criteria for validity and reliability. However, the use of simulations does rise a number of special issues that influence the validity, reliability and feasibility of the methods [38,39]. Standardized patients, computer programmes model-driven simulations and virtual reality devices were a representative of the work being done with simulation. As a group, these devices offer realistic challenges that enable the assessment of a variety of skills that are inaccessible to traditional methods [40]. Communication skills, the ability to take history from a patient or perform a physical examination, and the management of patients or experimental scenarios both overtime and in acute conditions are examples of the type of content to which these methods are well adapted [2]. The development of instructional guidelines and detailed scoring criteria are being researched, particularly for computer based simulations [41,42]. These efforts will most likely increase the integration of computer based simulators in medical education curricula and assessment. Issues of feasibility remain with many of these methods and there are underlying psychometric issues, such as how to equates scores, which need to be addressed in future research [43]. Proponents of using assessment with simulated patients have argued that there is a greater realism provided in clinical simulations than in paper-and-pencil exams specially in the affective and psychomotor domains [44-46]. In addition, the possibility exists that students can be evaluated on multiple components of clinical proficiency, such as observational skills, didactic knowledge, problem solving, clinical and technical skills, decision making and interacting with patients [44,47,48]. The tendency of this assessment method being realistic, having good construct validity and having acceptable reliability are also factors that support the use of simulated patient as evaluation tools [47,48]. Evaluation with simulated patients appears to be effective for assessing competencies that require direct interaction with the patient.

Tools for assessing Psychomotor domain

The model for assessment that was proposed by George Miller [49] is the most accepted one in medical education. ‘Miller Pyramid’ proposed clinical competence on multiple levels, ‘Knows’, ‘Knows How’, ‘Shows’ and ‘Does’. Choosing of assessment instrument is depending on using ‘Miller Pyramid’ [13]. However, the separation of clinical competence from one level to another is artificial, and knowledge and competence is a holistic entity [13]. Different levels of competence testing need a using of multi and variable assessment instruments. Assessment of psychomotor domain can be achieved by using a tools that assess the levels of ‘shows how’ and ‘does’ of Miller schema. Thus the tools that may be used to assess this domain are:

A. Long case:

A candidate assessed on one long case with oral presentation of information that he arrange them from history taking and physical examination. The student should answer questions related to the patient case (Non-standardized) and to perform physical examination under the observation of the examiner. In addition, he should perform an interpretation of the findings which lead to differential diagnosis and subsequently to professional diagnosis. This type of test provide a unique opportunity to assess the students task and interaction with real patient. However, this tool is with the limitation shown in Table 8.

The disadvantages of long case examination as we meet in TUCOM is that there is variation ($> 10\%$) in rater scale given to the student and a variation in the cases difficulty levels between the examinees. Noricini [50] reported that 61% of variability is due to errors in measurement and the reproducibility score drops to 0.24 with using single long case. This means that in our college as we use single long case, only 24% of variability of the score is due to actual performance of students. Long case difficulty is primarily a consequence of the fact that it is a single case examination [51], which may be influenced by chance of case selection (difficult or not; tired patient or well; helpful patient, e t c...). There is only a marginal effect on improving the reliability following standardization of patients, questions, and examiners [51]. In order to achieve a minimum acceptance of reliability of 0.85 required using of 10 cases and 200

minutes of testing time for each candidates [52] and this is impractical [13]. Significant improvement of long case can be achieved by increasing the number of cases, examiners, or assessed competencies [50].

B. Short case:

In TUCOM we use 3 short cases of non- standardized real patients with two examiners. Each examiner gave his mark individually and a sum for all cases. The short case tool are : Provides opportunity for assessment with real patients, using multiple cases, detailed assessment of clinical examination skills, and good construct validity [13]. However, it is with limitation such as inter- rater reliability is variable for the same examination and traditional short cases are less standardized than new forms (OSCE) [13]. In comparison to long cases, the short cases are better in discriminating between good and poorly performing students [53]. Short cases can be improved by using standardized multiple short cases that represent multiple competencies and a variety of problems [13].

C. Objective Structured Clinical Examination (OSCE).

Harden [54] describes OSCE as “ an approach to the assessment of clinical competence in which the components of competence are assessed in a planned or structured way “. By definition, the OSCE attempts to eliminate the subjectivity which besets most types of assessment and to focus on predetermined areas for testing. It concentrates primarily on the “ doing ” and learns the “ knowing ” to the more traditional forms of assessment.

The OSCE is a timed multi-station (usually 15-20) examination often using standardized patients to stimulate clinical scenarios [13]. The organizational aspect of the OSCE may vary somewhat but there are certain elements common to the examination. First, the clinical competencies and skills relevant to general practice that are considered worthy of assessment must be clearly defined. These competencies are then broken down into different components, for example history taking, data interpretation and simulated consultation using role play. The candidates moves round these different stations may vary in time lasting 4 to 10 minutes. The OSCE advantages and limitations [13,52,55,56,57] are shown in Table 9.

In the OSCE the roles are portrayed accurately and convincingly [58]. It is a specific example of the standardized patient examination used in medicine. The examination concentrate on hands-on- clinical behaviour where assessors ask students to show how to perform clinical tasks [59,60]. There is mixed feedback on the OSCE. The literature reports a high correlation between written and clinical examination formats, which dictates what is being measured [61]. Scores on written tests can predict students performance on the OSCE. Although, there appears to be a relationship between scores on written measures and the OSCE, its apparent face validity may make it useful when assessing the effectiveness of medical schools [62].

There is some support for the use of the OSCE in medical literature. It has been hypothesised that schools using the OSCE as part of their curriculum would have a particular advantages over students from schools that do not use it [62]. The examination concentrates on hands-on-clinical behaviour in which examiners ask students to “show how” to perform the select behaviour [13]. However, as originally proposed, the examination measures clinical skills in isolation and over short periods, which is not a valid representation of clinical reality at the end of a curriculum [63]. There is also support for the need to demonstrate of skills in isolation as the knowledge of skills, is important for the proper performance of skills [13]. To obtain a reliable measure of performance, an OSCE with 14 - 18 stations is recommended [31]. Furthermore, a fact that works in favour of test developer and examiners is the global rating in OSCE produce equivalent results as compared to checklist [51]. In OSCE the reliability is influenced by stations number rather than by time length of stations [64].

D. Objective Structured Practical Examination (OSPE):

Practical examination is an important component of evaluation in the medical curriculum. However, evaluation of students is not easy if the criteria of objectivity, uniformity, validity, reliability, and practicability have to be met [65]. At present practical exercise in basic sciences in most Iraqi medical colleges with the exception of TUCOM are conducted and evaluated in the traditional methods.

In this tool the students made to rotate through 15 -20 stations and undertook a variety of tasks, such as interpretation of graphs,

diagnosis of clinical conditions from photograph, identification of blood cells through microscope, reading packed cell volume, erythrocyte sedimentation rate, calculation of clearance, forced expiratory volume in 1 seconds, determination of serum lipid profile using spectrophotometer, determination of liver enzyme, etc. Each of these stations had questions carry 4 -5 points. Some stations including questions that emphasized the clinical relevance of basic sciences teaching. OSPE has the potential to motivate and influence learning approaches of students [66,67] and found to be reliable tool for assessing problem solving skills [66]. This assessment tool test dimensions of student performance because it tests both skills and knowledge [68,69].

OSPE has been suggested as an alternative for traditional instrument for the assessment of laboratory exercise [70]. In this the evaluation is structured in a way that all the educational objectives of an exercise can be assessed. Also in OSPE, questions and skills relevant for medical undergraduate training can be framed and tested keeping the educational objectives in mind. The advantages of OSPE include: objectivity and uniformity in the questions and marking of students [71]. In OSPE, I use three types of stations: The response stations, which consisted of short questions including graphs and diagrams requiring explanation; Procedural stations, where the candidate is expected to perform an assignment task and an examiner observed the student using a structured checklist; and stations for evaluation of communication skills.

E. Mini- Clinical Evaluation Exercise [Mini-CEX].

Mini-CEX is a rating scale developed by the American Board of Internal Medicine in the 1990s to assess seven core competencies of residents [13,72]: These include, medical interviewing skills, physical examination skills, clinical judgement, humanistic quality/professionalism, counselling skills, organization and efficacy, and overall clinical competence. Each competence is rated from 1-9 (1-3 unsatisfactory, 4-6 satisfactory, 7-9 superior). This tool is with advantages that include [13]: Direct observation of candidate performance, allows global evaluation of performance, good inter-rater reliability, practical and easy to use, and possible to customize to local context and needs. Although, the Mini-CEX has been the focus

of a number of research projects [72-75], a nontrivial limitation of this work has been the structure of the data collection. In Mini-CEX the trainees will be asked to undertake six observed encounters during 12 months with a different observer for each encounter. Each encounter lasts for about 15-25 minutes and for each encounter the evaluator records the complexity of the patients problem [low, moderate, high]; type of visit [new or return]; setting [ward, emergency room, clinic or ICU]; focus of the visit [data gathering, diagnosis, therapy or counselling]; time spent observing the encounter; and time spent in giving feedback [13]. Each of the encounters should represent a different clinical problem and trainees should sample from a wide range of problem groups within the year. Immediate feedback will be provided for after each encounter, by the observer rating the trainee. Trainers and trainees will need to identify and agree strengths area for development and an action plan for each encounter [76]. The limitation of Mini-CEX are unfamiliar and relatively new, faculty training is needed to improve reliability and assessment of all aspects of competences through single encounter is not possible [13]. However, Mini-CEX is with reliability and reproducibility of 0.73 and above [73]. In addition, it is helpful in discriminating different levels of performance [77]. Furthermore, improvement of reliability could be achieved by increasing the number of encounter and at least 4 -6 encounters are needed to reach acceptable reliability [73]. Mini- CEX is user and time – friendly and highly acceptable to both faculty and trainee [78]. The tool is suggested to be used in direct observation of students performance with real patients, feedback and formative assessment to the students and competency assessment [13]. Through being observed undertaking a number of cases, over a period of times with a number of different assessors, these individual brief encounters add up to provide a reliable measure of a trainee performance. Clinical skills can be difficult to assess reliability within the workplace and in the past such as assessment has therefore suboptimal. Mini-CEX provides a practical solution with significant learning benefit [76,79-84].

F. Checklist:

Initially in medical training education, assessment models that used checklists for the evaluation of student performance were

designed to meet the accreditation requirements. There are inherent problems with the use of checklists for assessment of clinical competence [85]. Checklist is described as ‘ ‘ commonly used in assessments to capture an observed behaviour or action of a student ‘ ‘ [13]. Five to seven Likert scale is used for rating in checklist (e.g. agree, somewhat agree, neutral, somewhat disagree, disagree). An example of checklist is shown in Table 10.

Checklist is a tool of examination with the advantages of easy to develop and capture actual action and performed behaviour [86]. Checklists are useful for evaluating any competency component that can be broken down into a specific behaviours or actions. The usefulness of checklists exists are documented by evidence for the evaluation interpersonal and communication skills and of patients care skills (history, physical examination, procedural skills). Self assessment of practice- based learning skills (Evidence based medicine) can be performed using checklists.. Checklists are most useful to provide feedback on performance because checklists can be tailored to assess detailed actions in performing a task [1,87]. However, checklists have a limitations that include [13]: Often casually developed and implemented, validity depends on the representative of items on checklists for the expected and desired competency, problem of inter-rater disagreement, single global rating scale using for evaluation [88]. In addition, students scores on checklists may tend to increase with increased expertise of the examiner [89]. Furthermore, student scores on checklists may also be affected by an evaluators perception of the reputation of an education program [90,91]. It is also possible that external raters on checklists may not be attend to the affective components of a clinical competency or its performance [90,92,93]. Checking boxes and marking scales may not contain details of an observed performance and provide no opportunity for the observer or faculty member to make judgments related to clinical performance [94,95]. However, when observers are trained to use checklists [96] consistent scores can be obtained and reliability in the range of 0.7 to 0.8 is reported. Performance score derived from checklists can discriminate between residents in different years of training [97]. Getting the agreement of experts improve the content validity of checklist [98].

G. Direct Observation of Procedural Skills (DOPS)

DOPS is the observation and evaluation of a procedural skill (practical or technical) performed by a trainee on a real patient. The procedural skills assessed using DOPS range from relatively simple and common procedures, such as venepuncture, through to more advanced surgical skills, such as endoscopic retrograde cholangiopancreatography. [99]. Evaluation by an experienced doctor is carried out using either a checklist of defined tasks, a global rating scale, or a combination of both. To properly define DOPS and distinguish it from several other assessments which bear close resemblance to DOPS a few key features of this assessment instrument should be pointed out. Firstly, DOPS involves the assessment of procedural skills rather than other clinical skills such as taking patient histories or performing physical examination [100]. Observation of these clinical skills would better be described as long cases with observation. Secondly, DOPS involves the evaluation of a specific patient encounter rather than a rating given based on observation over a period of time, as is the case in “supervisor” or “ward” evaluations. Thirdly, DOPS involves the performance of procedures on actual patients rather than cadavers, simulations or animal models [101]. Since students are observed performing procedures on patients in real settings, DOPS test at the “does” level [13]. Recent studies suggest that DOPS test is with benefit on learning procedure and student evaluation [81-84]. The advantages and limitations of DOPS are shown in Table 11 [13,102-108].

H. Clinical Work Sampling (CWS).

CWS is an in-trainee evaluation method [13]. In the clinical setting, real patients may not necessarily match the level of knowledge evaluated in the assessment. It is also possible that reliability is affected by the inability to separate student performance from other variables in the clinical situation [109,110]. In addition, the amount of faculty enthusiasm for doing observations can affect student scores on observation assessment [92,111]. CWS, Mini-CEX and DOPS address the issue of system and rater bias by collecting data on observed behaviour at the time of actual performance and by using multiple observers and occasions [13] and have the opportunity to provide feedback to the student and trainee. CWS form design takes into

account the content of patients encounters and different forms are used in different situations [13]. One of these forms the Admission Rating Forms to collect data on communication skills, physical examination skills, and global performance. Patient Rating Forms capture data on four domains: communication skills, collaboration skills, health advocacy skills and professionalism [112].

CWS is with the advantages of : direct of performance, authentic, multiple data source, consider different clinical situations, and include data from patients. However, it is with limitations such as , difficult to obtain data from patients, less well studied and relatively new [13]. The items evaluated in CWS are, Admission Rating Form, Ward Rating Form, Multidisciplinary Team Rating Form, and Patient Rating Form. The tool can be used in direct observation of performance in real clinical situations, feedback and competency assessment [13].

I. 360- Degree Evaluation.

360-degree evaluations consist of measurement tools completed by multiple people in a person's sphere of influence. Evaluators completing rating forms in a 360-degree evaluation usually are superiors, peers, subordinates, and patients and families [13].

Information regarding student's several topics performance (e.g., teamwork, communication, management skills, decision-making) are gathered using questionnaire or survey in most of 360-degree evaluation. Rating scales used in most 360-degree evaluations to assess how frequently a behavior is performed (e.g., a scale of 1 to 5, 1 meaning "never" and with 5 meaning "all the time"). Feedback is provided by summarizing ratings for all evaluators by topic and overall [13,31].

In 360- degree evaluation, the evaluators provide less lenient and more accurate ratings when the evaluation is intended to give formative feedback rather than summative evaluations. A professional behaviors, interpersonal and communication skills, some aspects of patient care and systems-based practice can be assessed using 360-degree evaluation. However, significant challenges are facing conduction of 360-degree-evaluations in clinical settings. The main challenges are: data gathering orchestration from a potentially large number of individuals that can be compiled and reported confidentially to the resident and constructing surveys that are

appropriate for use by all evaluators in the circle of influence. The feasibility of 360-degree-evaluation can be achieved by electronic implementation [113]. However, the tool provide assessment of actual action and behavior, assessment by multiple observers, provide evidence, and highly valued as a developmental tool.

J. Log Book

The student keeps a record of the patients seen or procedure performed either in a book or computer [13]. In TUCOM we use the book. This tool is with the following advantages [13,114]: students range of patient care and learning experiences was documented, direct students towards important objectives, and ensurance of students experience uniformity. However, it with limitations such as; “Accuracy of students reporting and faculty grading is difficult to ascertain, minimum number of procedures to be performed and cases to be seen is often set arbitrarily and is not validated against performance in the future [31], The number of procedures performed and patients seen does not necessarily correlates with competence achieved [31], and unlike portfolio, there is no scope for personal goal settings and reflections”[13].

K. Portfolio.

A portfolio is a collection of products prepared by the resident that provides evidence of learning and achievement related to a learning plan [29]. Typical portfolio consists of written documents, however, it may include other forms of information that are collected by residents such as photographs, audio- or video recordings. Additionally, statements about what has been learned, its application, remaining learning needs, and how they can be met may be included in portfolio. In graduate medical education, a portfolio might include a summary of the research literature reviewed when selecting a treatment option; a log of clinical procedures performed; a quality improvement project plan and report of results; a computer program that tracks patient care outcomes; ethical dilemmas faced and how they were handled; or a recording or transcript of counseling provided to patients [13,30]. Portfolio is a reflection of what the residents are learned. Both summative and formative residents evaluation can be performed using portfolio.

Portfolios are most useful for evaluating mastery of competencies that are difficult to evaluate in other ways such as use of scientific evidence in patient care, practice-based improvement, patient advocacy and professional behaviors. Morning reports, teaching experiences, patient rounds, research projects or individualized study are examples of learning experiences that lend themselves to using portfolios to assess residents [30,31,115]. Portfolios acceptance in graduate medical education varies according to preferred learning style. Some physicians and residents have found that by maintaining portfolios credit was allowed for some activities that otherwise would have gone undone or un-noticed. Yet, for others, the time and commitment necessary to create and maintain a portfolio is too great relative to the return [13,116].

Portfolio assessments reproducibility are feasible when there is agreement on standards and criteria for portfolio contents. When portfolio assessments have been used to evaluate an educational program, the portfolio products or documentation have been found to be sufficient for program evaluation but are not always appropriate to use in assessing individual students for decisions about promotion to the next grade. However, standard criteria are not necessarily desirable and may be counter-productive when the portfolio purpose is to demonstrate individual learning gains relative to individual goals. The validity of portfolio assessment is determined by the extent to which the products or documentation included in a portfolio demonstrates mastery of expected learning [13,116].

Table 1. Dave's Taxonomy [5].

<i>Level</i>	<i>Category</i>	<i>Behaviours description</i>	<i>Examples of activity or demonstration and evidence to be measured</i>	<i>Some verbs</i>
1	<i>Imitation</i>	<i>Student observing and copying some one else</i>	<i>Watch teacher or trainer and repeat action, process and activity</i>	<i>Copy, follow, replicate, repeat, adhere</i>
2	<i>Manipulation</i>	<i>Reproduce activity from instruction or memory</i>	<i>Carry out task from written or verbal instruction</i>	<i>Re-create, build, perform, execute, implement</i>
3	<i>Precision</i>	<i>Execute skill reliably, independent of help</i>	<i>Perform a task or activity with expertise and to high quality without assistance or instruction; able to demonstrate an activity to other learners.</i>	<i>Demonstrate, complete, show, perfect, calibrate, control</i>
4	<i>Articulation</i>	<i>Two or more skills combined, sequenced and performed consistently. Adapt and integrate expertise to satisfy a non-standard objective</i>	<i>Relate and combine associated activities to develop methods to meet varying, novel requirements</i>	<i>Construct, solve, combine, coordinate, integrate, adapt, develop, formulate, modify, master</i>
5	<i>Naturalization</i>	<i>Two or more skills combined, sequenced and performed consistently and with ease. Automated, unconscious mastery of activity .</i>	<i>Define aim, approach and strategy for use of activities to meet strategic need</i>	<i>Design, specify, manage, invent, project-manage.</i>

Table 2 Simpson's Taxonomy.[6]

Level	Category	Description	Examples of activity or demonstration and evidence to be measured	Some verbs
1	Perception	Awareness-sensory cues guide motor activity	Use and/or selection of senses to absorb data for guiding movement	Recognize, distinguish, notice, touch, hear, feel,
2	Set	Readiness	Mental, physical or emotional preparation before experience or task	Arrange, prepare, get set
3	Guided Response	Attempt	Imitate or follow instruction, trial and error	Imitate, copy, follow, try
4	Mechanism	Basic proficiency-intermediate stage in learning a physical skills.	Competently respond to stimulus for action	Make, perform, shape, complete
5	Complex Overt Response	Expert proficiency	Execute a complex process with expertise	Coordinate, fix, demonstrate
6	Adaptation	Adaptable proficiency-movements can be modified for a special situation	Alter response to reliably meet varying challenges	Adjust, integrate, solve
7	Origination	Create proficiency	Develop and execute new integrated responses and activities	Design, formulate, modify, re-design, trouble-shoot

Table 3. Harrow's Taxonomy.[7]

<i>Level</i>	<i>Category</i>	<i>Description</i>	<i>Examples of activity or demonstration and evidence to be measured</i>	<i>Some verbs</i>
<i>1</i>	<i>Reflex Movement</i>	<i>Involuntary reaction</i>	<i>Respond physically instinctively</i>	<i>React, respond</i>
<i>2</i>	<i>Basic Fundamental Movements</i>	<i>Basic simple movement</i>	<i>Alter position, move, perform simple action</i>	<i>Grasp, walk, stand, throw</i>
<i>3</i>	<i>Perceptual Abilities</i>	<i>Basic response</i>	<i>Use than one ability in response to different sensory perceptions</i>	<i>Catch, write, explore, distinguish using senses</i>
<i>4</i>	<i>Physical Abilities</i>	<i>Fitness</i>	<i>Develop strength, endurance, agility, control</i>	<i>Endure, maintain, repeat, increase, improve, exceed</i>
<i>5</i>	<i>Skilled Movements</i>	<i>Complex operations</i>	<i>Execute and adapt advanced, integrate movements</i>	<i>Drive, build, juggle, play a musical instrument, craft</i>
<i>6</i>	<i>Non-discursive Communication</i>	<i>Meaningfully expressive activity or output</i>	<i>Activity expresses meaningful interpretation-body language</i>	<i>Express and convey feeling and meaning through movement and actions</i>

Table 4. Combination list of taxonomies.

Level	Category	Definition	Example
1	Observing	Active mental attending of a physical event	The learner watches a more experienced person . Other mental activity, such as reading may be a part of the observation process
2	Imitating	Attempted copying of a physical behaviour	The first steps in learning a skill. The learner is observed and given direction and feedback on performance. Movement is not automatic or smooth
3	Practicing	Trying a specific physical activity over and over	The skill is repeated over and over. The entire sequence is performed repeatedly. Movement is moving towards becoming automatic and smooth
4	Adapting	Fine tuning, Making minor adjustments in the physical activity in order to perfect it	The skill is perfected > A mentor or a coach is often needed to provide an outside perspective on how to improve or adjust as needed for the situation

Table 5. Writing Objectives: Key Verbs. Psychomotor (Doing or Skills) Domain [TUCOM Documents]

Absorb;	Demonstrate	Freeze;	Record;
Add;	Describe	Grade	Release
Adsorb;	Design;	Grasp;	Remove;
Adjust	Dialyze	Grind	Replace
Aliquot;	Differentiate	Group;	Re-suspend
Apply;	Dilute;	Guide	Re-test;
Aspirate;	Discard	Handle;	Rinse;
Assemble;	Dismantle	Heat	Roll;
Arrange	Dispense	Hemolyze;	Rotate;
Analyze	Dispose;	Identify	Save;
Appraise	Dissect	Illustrate;	Scan;
Attach	Dissolve;	Incubate	Score;
Assess	Drain;	Inject;	Screen;
Balance;	Draw;	Input	Seal; Select;
Bind;	Dry;	Insert;	Sensitize
Blend;	Determine;	Invert	Separate;
Build;	Define	Investigate	Set
Calculate;	Develop	Isolate;	Sever;
Calibrate;	Defend	Label	Shake
Centrifuge	Discriminate	Locate;	Sharpen;
Change;	Duplicate	Localize	Ship
Choose;	Elute	Lyse;	Siphon;
Classify;	Employ;	Maintain	Spin
Clean;	Estimate	Make;	Spread;
Collate;	Evacuate;	Maneuver	Squeeze
Collect;	Evaluate	Manipulate;	Stain;
Combine;	Examine	Mark	Start
Connect;	Expel;	Macerate;	Standardize
Construct;	Fasten	Measure	Stick;
Critique	Fill;	Mix;	Stir
Control;	Filter;	Moisten	Stop;
Confirm;	Fractionate	Mount;	Stopper
Connect;	Frame;	Observe	Store;
Contrast ;	Pour;	Obtain;	Suspend

Control;	Prepare;	Open;	Take;
Cool;	Press;	Operate;	Test
Correct;	Process;	Pack;	Thaw;
Count;	Produce;	Palpate;	Thread
Create;	Program;	Participate	Tilt;
Crush;	Pull;	Perform;	Time
Cut;	Puncture;	Pick;	Tip;
Clarify;	Push;	Pipit;	Titrate
Compute	Predict	Place;	Position
Conclude	Paraphrase	Plate;	Duplicate
Decant;	Read;	Plot	Express
Utilize;	Trim;	Turn;	Extend
View;	Touch;	Type	Divide
Warm;	Transfer	Rate	Infer
Wash;	Troubleshoot	Generate	Interpolate
Watch;	Label	Employ	Extrapolate
Weigh;	Identify	Diagram	Rephrase
Withdraw;	Sketch	Integrate	Reorder
Wipe	Show	Inventory	Rearrange
Use;	Write	Execute	Restate
Wrap	Outline	Break	Transform
Value	Review	Propose	Tell
List	Associate	Synthesize	Invent
Label	Contrast	Perform	Order
Recall	Implement	Specify	Sketch
Quote	Structure	Score	Practice
Tabulate	Criticize	Judge	State
Tell	Measure	Revise	Create
Reproduce	Devise	Relate	Recommend
Retrieve	Formulate	Support	Hypothesize

Table 6 . Strengths and limitations of performance assessments.
[1]

Strengths	<ol style="list-style-type: none"> 1. <i>Can evaluate complex learning outcomes and skills that can not be evaluated with traditional paper-and-pencil tests.</i> 2. <i>Provides a more natural , direct, and complete evaluation of some types of reasoning, oral and physical skills.</i> 3. <i>Provides greater motivation for students by clarifying goals and making learning more meaningful.</i> 4. <i>Encourages the application of learning to real life situations</i>
Limitations	<ol style="list-style-type: none"> 1. <i>Requires considerable time and efforts to perform.</i> 2. <i>Judgement and scoring performance is subjective, burdensome, and typically has low reliability.</i> 3. <i>Evaluation must frequently be done individually , rather than in groups</i>

Table 7. Improving performance assessments [1]

No	Description
1	<i>Specify the intended performance outcomes in observable terms and describe the use to be made of the results</i>
2	<i>Limit the observable dimensions of the performance to a reasonable number</i>
3	<i>Provide clear , definite criteria for judging the procedure or product</i>
4	<i>Select the performance setting that provides the most relevant and realistic situation</i>
5	<i>If a structured performance situation is used , provide clear and complete instructions</i>
6	<i>Be as objective as possible in observing , judging, and recording the performance.</i>
7	<i>Observe the performance under various conditions and use multiple observations whenever possible</i>
8	<i>Make a record as soon as possible after an observation</i>
9	<i>Use evaluation forms that are clear , relevant, and easy to use</i>
10	<i>Use scoring procedure that is appropriate for the use to be made of the results</i>
11	<i>Inform students of the method and criteria to be used in evaluating the performance</i>
12	<i>Supplement and verify performance assessments with other evidence of achievement</i>

Table 8 . Limitations of long case [13]

<i>No</i>	<i>Limitation type</i>
<i>1</i>	<i>Serious doubts about reliability and consistency.</i>
<i>2</i>	<i>Poor content validity as only 1-2 cases are tested [in TUCOM]</i>
<i>3</i>	<i>Generalizability across other competencies is poor</i>
<i>4</i>	<i>Assessment relies on candidates presentation, represent assessment of “Knows how” – a lower level competency rather than “Show how”.</i>

Table 9. OSCE advantages and limitations [13,52,55,56,57]

<i>Advantages</i>	<p><i>*An effective alternative to unstructured short cases, since OSCE allows examiners to determine in advance those competencies to be tested and subsequently to design the examination to test these competencies.</i></p> <p><i>* Allows a sample of wide range of skills and competencies across a number of stations.</i></p> <p><i>* Allows standardization of cases .</i></p> <p><i>* Greater reliability and marking.</i></p> <p><i>* A wide range of skills and behaviours may be looked at and assessed under varying degrees of difficulty and complexity.</i></p>
<i>Limitations</i>	<p><i>*OSCE tends to compartmentalize a candidate skills and knowledge. This is particularly important criticism as perhaps the sum of individual parts does not always equal the whole.</i></p> <p><i>* The obvious problems relate to time, labour intensive, and expensive</i></p> <p><i>* Communication assessment, especially the attitudes, is difficult, as these skills are case specific and have poor generalizability.</i></p> <p><i>* Validity of the test is compromised if a complex skill, Is fragmented into multiple minor tasks.</i></p> <p><i>* OSCE relies on task – specific checklists which assumes that physician- patient interaction can be described as a list of actions.</i></p>

Table 10. Checklist for laboratory experiment performance

Skill	1	2	3	4	5
Following instructions					
Selecting appropriate apparatus					
Using appropriate manipulative skills					
Conducting experiments safely					
Making accurate observations and measurements					
Recording results and observations accurately and clearly					
Presenting experimental results clearly					
Interpreting and evaluating observations and data					
Drawing conclusions and making generalizations from experiments					
Planning and organizing experimental investigations					
Identifying problems					
Evaluating methods and suggesting improvements					
Communicating plans and results					
Identifying and controlling variables					

Scores:

1= Unacceptable.

2= Marginally satisfactory

3= Satisfactory

4= Good

5= Excellent

Table 11. Advantages and limitations of DOPS [13,112-118].

Advantages	<ul style="list-style-type: none"> • <i>Direct observation of procedural skills.</i> • <i>Allows global evaluation.</i> • <i>Practical and easy to use.</i> • <i>Possible to customize to local contexts and needs.</i>
Limitations	<ul style="list-style-type: none"> • <i>Relatively new and unfamiliar.</i> • <i>Faculty training is needed.</i> • <i>It is not possible to assess all aspects of competencies through a single encounter.</i> • <i>For technical procedure there is a need to have expert observer</i>