

# THE ROLE OF WORKPLACE-BASED ASSESSMENTS IN ASSESSING THE COMPETENCIES OF RESIDENTS

## Abstract

Workplace-Based Assessment (WPBA) represents a pivotal shift in medical education, emphasizing real-time evaluation of residents' clinical competence within authentic patient care environments. Unlike traditional assessment tools that focus on knowledge or controlled simulations, WPBAs aim to assess the "Does" level of Miller's pyramid—direct performance in real-world clinical settings. This model of assessment enhances educational validity by capturing the integration of knowledge, skills, and attitudes required for effective patient care.

WPBA tools—such as Mini-Clinical Evaluation Exercises (Mini-CEX), Direct Observation of Procedural Skills (DOPS), Case-Based Discussions (CbDs), and Multi-Source Feedback (MSF)—facilitate formative feedback, promote reflective practice, and support competency-based medical education (CBME). These tools are especially valuable in fostering feedback-rich environments and tracking learner progression over time. Studies from global and Indian settings confirm WPBA's acceptability, educational impact, and role in improving clinical performance.

However, challenges remain. Faculty training, time constraints, assessor variability, and the tension between formative and summative use affect the reliability and acceptance of WPBA. Additionally, WPBA's effectiveness depends heavily on high-quality feedback, observation-based assessment, and clearly defined competencies. Entrustment-based scales and tools such as EPA-IC in surgical specialties have enhanced the alignment of assessment with clinical expectations.

To strengthen WPBA's role in medical training, systemic strategies are needed: improved faculty development, clearer assessment frameworks, and integration of feedback into daily practice. Despite

## Author

**Dr. Pabithadevi**

Professor and Head

Department of General Surgery

Member, Medical Education Unit

Tirunelveli Medical College

MHPE scholar, Sri Balaji Vidyapeeth,

Pondicherry, India.

limitations, WPBAs represent a robust method to assess and guide the development of clinical competence, professionalism, and communication in postgraduate medical education, aligning assessment closely with clinical reality.

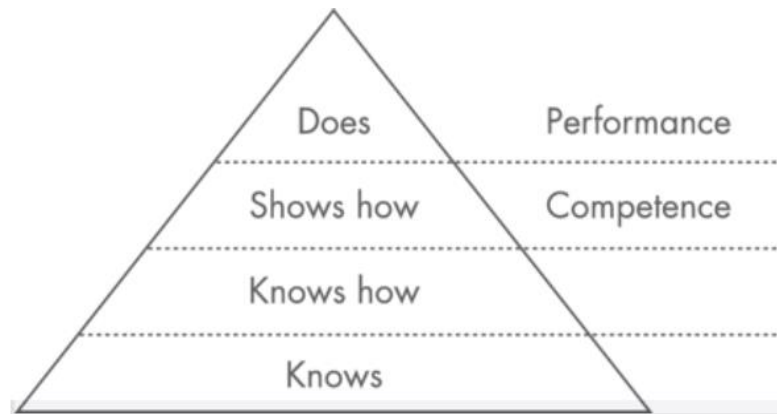
## **I. IMPORTANCE AND RELEVANCE**

The medical undergraduate trainee or the postgraduate resident must undergo training in real life setting under supervision to attain the level of competent practitioner from novice. Assessment has been changed a lot in medical education in last five decades. The assessment needs to evaluate a broader range of clinical situations and should include more complex issues and emergency care, various system diseases and practical skills. The focus of the assessment should be on more complex integrated skills rather than a single component of competency. Also, the results of evaluation program should support the curriculum design of educational architecture. It is hard to design good quality written or performance-based assessment which can assess the competency of the trainees. Additionally, there is a requirement to address complex material and abilities that are challenging to simulate at all times. However, routinely, during training, there are interactions between the treating physicians, trainees, and patients where the clinical content will be used as a foundation for evaluation. Besides, the clinical educators who are available can serve as the assessors.

WPBAs are designed to provide opportunities for structured observation and feedback in a low-stakes setting, as well as contribute to the accomplishment of competency-based learning objectives (assessment for learning)<sup>[1]</sup>. The focus must be on maximising the utility of WBAs as formative assessment, as outlined in Graduate Medical Education, by linking clinical competencies with the curricula of graduate medical education (GME)<sup>[2]</sup>.

Workplace based assessment is the evaluation of trainee's professional knowledge, skill and attitude in the workplace i.e. in the real life setting. WBPA has high content validity. It is acceptable, reliable, cost effective and has an educational impact. WBPA are included in the educational curriculum because the learning objectives, teaching learning methods, assessment and feedback are aligned. The standard which is expected to be achieved at the completion of training is used for assessment of learners. WBPA provides evidence of learning and achievement and also points out the areas which needs improvement and addressing the ways to improve them.

Miller's pyramid is a helpful tool for evaluating clinical performance.



**Table 1:** Miller's Pyramid

The knowledge basis ('Knows'), as determined by easy knowledge assessments, appropriately forms the basis of the pyramid. The "Knows how" level, assesses comprehension and knowledge application and can be done through short essay questions, patient management problems, etc. The third stage, "Shows how," can be measured using tools like the OSCE. However, there is little correlation between how doctors perform during controlled examinations and how they perform in real-world settings. Thus, it is necessary to evaluate at the "Does" level, which is the highest level possible <sup>[3]</sup>. WBA evaluates the effective application of competencies in real-world contexts. The WPBA is compatible with Miller's pyramid's highest level (Level 4: "Does") and has the capacity to evaluate at all four levels.

A faculty observes the trainee in a specific patient encounter, case record or a specific procedure and assesses it. It shows that the performance of the trainee has been observed by the assessor and focused on reaching the decision about the performance. However, in medical education, the performance of the physician is patient or situation specific and the performance of one event cannot predict the same performance in other tasks also. To provide a reliable measure of the trainee's performance, several occurrences must be evaluated.

But how it differs from the routine observations, the assessors have made over the time by observing the performance of the trainees? The assessors, sometimes give evaluation of certain aspects of performance, they didn't observe. According to Pulito et al, the faculties mainly observe the cognitive skills and professionalism rather than other aspects of competency <sup>[4]</sup>. Silver et al also found out that The faculties often evaluate the ability of medical knowledge and interpersonal skills without discriminating between other aspects of competencies.

Depending on the purpose of assessment, the number of encounters of each WPBA may vary. An alternative to measuring reliability is standard error of measurement.

It is possible to create a 95% confidence interval throughout the results using scores. In a study of mini-CEX, if we take the scale as 1-3 is unsatisfactory, 4-6 is satisfactory, and 7-9 is expert, then the 95% confidence interval for the rating of clinical competence, was  $\pm 1.2$

after two encounters,  $\pm 0.8$  after four encounters, and it decreases further by increasing the encounters. The width of the confidence interval narrows and the frequency of good decisions rises as encounters are added. For trainees with an average rating of 6 or more, two encounters are enough to identify which trainees are unsatisfactory <sup>[5]</sup>. The trainees with borderline scores will require more interactions, and since each interaction is followed by feedback, the trainees who require more extensive educational interventions will receive them.

The comprehensive nature of WBAs makes them a equitable assessment format. Educational collaborators preferred WBAs with clear definitions and structure since they could aid assessors in making decisions, especially when it comes to lowering cognitive load and subjectivity and creating feedback. WBAs were thought to be particularly adept for charting trainees' development over time. Negative opinions of WBAs were caused by their improper use, which included failing to conduct direct observation, filling out retroactive forms, and providing insufficient feedback <sup>[6]</sup>. To provide high-quality patient care, CBME requires specific competencies and goals that go beyond clinical reasoning and medical knowledge <sup>[7]</sup>. Implementing CBME could provide learners with guidance and direction as well as accountability and transparency to patients and the broader public. According to the International CBME Collaborators, competency defines a quantitative and observable medical skill, whereas competence refers to a range of medical abilities <sup>[8]</sup>. Medical educators have repeatedly attempted to define competencies for CBME.

## History of WPBA

In the past, too much attention has been placed on gauging a student or trainee's ability to pass a test and not enough on whether they can execute to the standards required of medical practitioners. For many years, traditional clinical exams like the Objective Structured Clinical Examinations (OSCEs), which Ronald Harden invented in Dundee <sup>[18]</sup>, have been used extensively in a variety of educational settings. However, these evaluations have their limits. The doctor-patient encounter is frequently "deconstructed" by stations by having students do discrete parts of clinical encounters, and the kinds of patient problems that can be utilised are limited by the kinds of cases that can be simulated. The centre of attention of medical education is shifting quickly away from achieving a specific score on difficult exams and towards accumulating proof of clinical proficiency and professional conduct every day in the job. In order to evaluate workplace-based learning programmes, on-the-job workplace-based assessments (WPBA) have been established.

WPBA is the assessment of performance of a trainee based on their actual job. The major objective of WBA is to examine the facets of actual real time performance that are unsuitable for evaluation via a distance assessment of competence. By providing trainees with insightful feedback, it is perfect for the objective of encouraging learning (evaluation for learning). The same process can be applied by trainees during reflective practise to evaluate themselves. The evaluations assist the supervisor in tracking a learner's development during a placement <sup>[5]</sup>.

The Mini-Clinical Evaluation Exercise, a workplace-based assessment instrument expressly created to provide feedback following an observation of a physician-patient clinical encounter, was published by Norcini et al. in 1995<sup>[20]</sup>. Since then, more than 50 assessment tools have been created to assess particular facets of clinical practise, including instruments to

evaluate clinical/practical skills, clinical reasoning, and clinical behaviours. In order to explore the viability of using thorough workplace assessments across all medical specialties in the UK, Wilkinson et al. undertook a feasibility study. Between 2003 and 2004, from 17 specialties 230 SpRs participated at 58 UK hospitals. In this study, it took 25 minutes to finish the mini-CEX (including feedback) <sup>[21]</sup>. The DOPS required one-third more time for feedback than the length of the procedure being evaluated. Each Rater needed an average of 6 minutes to complete their MSF form. They also discovered that the techniques can reliably distinguish between the performances of different doctors and are practicable to use. They might be suitable for evaluating the performance of other doctor grades and specialties in the workplace with some adaptation.

An assessment should have high reliability, validity, educational impact and acceptable to assessors and trainees and feasible.

**Validity:** It is the extent to which an assessment instrument measures what is supposed to measure. It is concerned with whether the appropriate evaluations are being made, how they are being made, and whether learning is being positively impacted. There are different varieties of validity. They are content, face, construct, predictive, and consequential validity.

**Content Validity:** If every component of an assessment aligns with the competencies (knowledge, skills, or behaviors) that it is intended to evaluate, then the assessment has content validity.

**Face Validity:** Content validity and face validity are connected. It is explained from the assessor's point of view. The evaluation has good face validity if the assessor believes it to have a high level of validity.

**Construct Validity:** The degree to which an assessment's multiple facets, including the various subcomponents, test the underlying professional constructs is termed as construct validity.

**Predictive Validity:** The degree to which an evaluation predicts anticipated results is referred to as predictive validity.

**Consequential Validity (Educational Impact):** The effect of an evaluation on learning is indicated as consequential validity (educational impact) <sup>[5]</sup>.

Rasmussen brought forward that the first stage in practical learning is the acquisition of skills and that should be competent in these skills before the full knowledge relating to the skill has been acquired. The skills can then be applied by students using a series of rules. He theorises that with time the practical experience improves and is augmented with knowledge. It would be ideal for the student to finally advance to the highest possible level of "knowledge-based practice."

Current literature discusses Rasmussen's theories, and Long cites this notion as the foundation for competency-based training in medicine, which was introduced more recently. Rasmussen's concept is more relevant to procedural specialties than to medical practice, where practical techniques are used seldom<sup>[22]</sup>.

Different models have been proposed out in order to study clinical medicine. Initially, Dreyfus and Dreyfus created a five-stage model to explain how a pilot's abilities and expertise grow. In a recent study, they try to broaden this model to include certain elements of "performance" and the development of clinical competencies in medicine. Dreyfus stages applied to clinical medicine are as follows.

Stage 1—novice (medical student)

Stage 2—advanced beginner (house officer or SHO)

Stage 3—competent (registrar)

Stage 4—proficient (newly appointed consultant)

Stage 5—expert (mid-career physician)

According to this, student Learns basics of clinical examination, clinical skills and slowly learns how to use abilities in specific clinical scenarios that depend more and more on the circumstances—such as hospital admissions, rounds, etc.—allowing for experience-based learning. And learns treatment approach to patient's condition under supervision and finally develops routines to streamline patient care. Manages multiple situations in a thoughtful way which is intellectually and emotionally absorbing and integrates mastered skills with personal style.

Miller proposed a more practical approach in his famous triangle. According to him, there are four stages of development: "knows, knows how, shows how, and does." These are the behavioral and cognitive phases that a person undergoes once they progress from learning something to actually doing it. Clinical performance may be influenced by numerous additional aspects in clinical practice.

Knowing and showing does not mean that a physician will perform in a similar way in real practice. To address these concerns, "The Cambridge model" is proposed by Rethans and colleagues as a modification to Miller's triangle<sup>[23]</sup>. In addition to competence, they aimed to describe the elements determining "performance" by expanding Miller's triangle. They determine that performance is what a doctor exhibits in actual clinical practice, whereas competence is what a doctor exhibits in a test setting. Competence is the foundation of performance, which also takes into account other factors that may affect an individual's final result. They created a triangle that was altered. This model recognized the necessity to evaluate doctors in training's global competency or performance in a realistic manner in addition to their knowledge and practical skills. They ascertain that competence is what a doctor demonstrates in a test situation, but that performance is what a doctor demonstrates in real life clinical practice. Performance builds upon competence but also encompasses other influences on one's eventual performance. They produced a modified triangle. This model acknowledged that in addition to assessing knowledge and practical skill we need to assess the global competence or performance of doctors in training in as realistic a way as possible<sup>[23]</sup>.

Assessment of Miller's pyramid stages 1 and 2 is performed by using traditional assessment tools including written and oral tests. Multiple-choice questions (MCQs), best-of-five questions, extended matching, and short answer questions are used in written tests to evaluate stages 1 and 2 of Miller's pyramid in medical, surgical, and most secondary care specialties.

However, knowledge and skills might not always translate into clinical performance or professional effectiveness. It is more difficult to evaluate Miller's triangle levels three and four. Objective structured clinical examinations (OSCEs) and practical clinical examinations are presently used to evaluate level 3 in medical disciplines. A number of alternative techniques, such as the mini-clinical evaluation exercise (mini-CEX), have been developed and proven effective. As long as the activities' content is suited to the exam's objectives, this kind of evaluation can be utilized both formatively and summatively. The number of stations and abilities assessed to provide a comprehensive picture of the trainees' clinical competence level determines the validity of an OSCE and other evaluation instruments. Testing the instrument's validity, reliability, and utility is crucial when creating an OSCE or other clinical assessment tool, particularly if the proposed exam is a high-stakes exam that will be conducted at multiple locations.

The OSCE was deemed a valuable instrument by Sloan et al. for evaluating the knowledge gap [24]. Novack et al. explained the creation of an online OSCE system for teleconference-based clinical skill evaluation<sup>[25]</sup>. They conducted the same clinical evaluation at multiple locations at the same time using videoconferencing and the internet. Students found it challenging to avoid direct contact with the patients, and also noted technical issues with the system. Other studies have described computer systems used to assess clinical competency<sup>[26]</sup>, and these technologies may eventually affect clinical education assessment systems. The literature describes other computer-based clinical assessment techniques. One approach is called "Primum E Simulations," where applicants use a computer to navigate a clinical situation while being solely informed by changing facts. Candidates are graded based on the tests and treatments they select, and the assessor looks for evidence of clinical reasoning<sup>[27]</sup>.

Level 4 of Miller's triangle, "does," or performance, is the most challenging aspect of clinical competence to assess. Even if we have the means to evaluate performance in a test setting, this does not always reflect what doctors actually accomplish in their day-to-day work. To guarantee that clinical abilities are evaluated effectively, it is crucial to examine aspiring doctors up close. This kind of evaluation can be expensive and time-consuming. To demonstrate a variety of clinical abilities, the Royal College of General Practitioners uses a system that involves seeing a collection of filmed consultations that the trainee submits<sup>[22]</sup>.

A number of authors support the use of standardized patients in these evaluations; nonetheless, patient training, the creation of legitimate instances, and scoring systems all require careful consideration. Students in the US go through ten simulated patient observations as part of the clinical skills evaluation needed to be certified as foreign medical graduates<sup>[28]</sup>. The mini-CEX, which was created in the 1990s based on the classic CEX or long case evaluation, is an alternative method used by the American Board of Internal Medicine to evaluate residents' clinical competence. It assigns scores to the clinical abilities that residents exhibit during patient contacts. Chief residents or supervising physicians score the tests, which are based on actual clinical scenarios<sup>[28]</sup>. In the mini-CEX, an assessor watches a student perform throughout a typical clinical interaction that lasts 15 to 20 minutes. The trainee is then evaluated using a 9-point scale on their proficiency in medical interviewing, physical examination, professionalism, clinical judgement, counselling, organisation, efficiency, and overall competence. This is followed by immediate feedback at the conclusion of the consultation, which lasts between five and fifteen minutes. A number of researchers and institutions agree with the mini-CEX's validity and reliability in assessing the

clinical competence of trainees. The mini-CEX is also recognised as an effective teaching and learning tool for increasing clinical achievement among trainees because of the linked feedback component.

Recently, for medical staff, 360-degree appraisals have been driven by structured or open-ended questionnaires answered by team members. The Royal College of Physicians is testing this method to evaluate specialist registrars in several disciplines <sup>[22]</sup>. Additionally, personal journals and portfolios can be used as effective tools to gauge attitude to a certain extent.

The direct observation of trainee's skill, knowledge and attitude became particularly important since the Accreditation Council for Graduate Medical Education (ACGME) created the Next Accreditation System (also known as the Mile- stones) in 2012.

It is probably required to combine the several assessment techniques mentioned above in order to create the most accurate assessment of global clinical competence. The new competency-based curriculum for SHOs in anesthesia has been evaluated by the Royal College of Anaesthetists using this method. Clinical skills, attitudes, and behaviors are evaluated in a workplace assessment conducted by supervising consultants, and SHOs' knowledge is evaluated through a series of written exams.

In earlier times, logbooks and the judgment of educational supervisors have been used to evaluate efficiency in practical tasks. Directly observed performance is probably more valid and reliable than the old logbook-based system, according to the Royal College of Physicians, which created the DOPPS instruments.

Multi Source Feedback has primarily been utilized in business and industry to evaluate performance and give trainees feedback. This approach evaluates both the medical and non-medical components of performance using information from peers' questionnaires. The mini peer assessment tool, often known as Mini-PAT, is a multi-source feedback tool that compares trainees' self-assessed performance with the opinions of a variety of clinical peers. The educational supervisor then provides the trainee with the assessors' ratings and free text comments.

One of the main objectives of formative assessment is to develop self-regulated learners who can independently recognise their own learning requirements, develop a plan to meet those needs, and, most crucially, self-monitor their progress.

The provision of quality feedback is essential in ensuring that an assessment is formative. Despite this, it is yet unknown how these WBAs will be utilised, in part because of the different applications that have been made for the tools. The constant "tick-box" technique for completion has been highlighted in research ever since, worries about relevance and accountability, and problems with implementation and acceptability <sup>[29]</sup>.

Some details of individual WPBA methods are given below.

### **Types of Workplace Based Assessment Tools**

Workplace-based assessment tools include:

- case-based discussion (CbD)



- directly observed practical skills (DOPS)
- mini clinical evaluation exercise (Mini-CEX)
- evaluation of clinical events (ECE)
- multi-source feedback (MSF)
- Mini Peer Assessment Tool (mPAT)
- Portfolio
- Entrustment aligned Pathology assessment instrument for intraoperative consultations (EPA-IC)
- Workplace based assessment tools in Emergency department

## II. DEFINITION OF WORKPLACE-BASED ASSESSMENT TOOLS

### Case-Based Discussion (CbD)

Case-based discussion (CbD) is a way for trainees to present and discuss their cases with their facilitator throughout their training and obtain systematic and structured feedback from the assessor. It is used to assess the decision-making skills based on their clinical and laboratory findings. The trainee must discuss why they have planned that line of management. The discussion based on the documented involvement of the trainee in medical notes or reports. The trainee themselves chooses the cases, the time, and the assessor. A case is selected with curriculum objectives in mind a few days before to the assessment, and it is then reviewed using targeted questions intended to elicit answers that would reveal knowledge, abilities, attitudes, and behaviours pertinent to those domains. After the discussion, the assessor assigns a quality score to the performance and provides helpful criticism. Trainees are evaluated six times a year on average.

Instead of evaluating what the trainees might have done, CbD evaluates what they actually did. And this is the key distinction between the Objective Structured Clinical Examination (OSCE) and CbD, which assesses the trainee's performance under examination conditions. CbD has been demonstrated to have significant face and content validity with good levels of reliability and validity can be achieved with assessor training <sup>[9]</sup>. Due to the inherent nature of CbD, evaluations of trainees' applied knowledge, clinical reasoning, and decision-making must be conducted mostly through interactions or discussions involving the doctors' own patients, or cases. The ability to identify difficulties, manage a complex case within the range of possibilities provided, choose a course of action, explain the course of action, and reflect on the results are just a few examples of the wide range of comprehensive, balanced, and justified solutions that CbD may explore in complex scenarios.

### Direct Observation of Practical Skills (DOPS)

The method called DOPS was created especially to evaluate trainees' efficiency in the daily practical tasks they perform as part of their training. The Royal College of Physicians first created and assessed the DOPS tool. The procedure may be a laboratory technique or a clinical procedure. The trainees decide on the process and schedule. However, the assessor must concur that the procedure is appropriate. The common procedures for which DOPS can be used are insertion of an IV-line, endo tracheal intubation, urinary bladder catheterisation and Ryle's tube insertion. Usually, the encounters are for 15 minutes with 5 minutes feedback. After the encounters, the faculty will rate the trainee's performance and will give

the educational feedback. The strength of the trainee and area of improvement will be documented. A standard format is used, and the feedback and comments are given immediately. The trainees will be evaluated several times by different examiners. Consultants, senior specialist registrars, associate specialists, and general practitioners are among the assessors.

The results of the global ratings can be reliable, according to studies. In an objective structured assessment of procedural skills, Goff et al. showed that two assessors' evaluations of occurrence, quality, and fitness allowed them to differentiate between the various training levels<sup>[73]</sup>. Marriott et al. found that using DOPS to assess the training skills in the operating theatre had good validity, reliability and acceptability<sup>[68]</sup>

### **CSR (Chart- stimulated recall)**

Maatsch created it for the American Board of Emergency Medicine to use.[74] Case based Discussion which is used in the foundation programme is a variation of CSR. In these settings, the trainee must select two case records from the patients they have seen recently. The assessor chooses one and works with the learner to study any one aspect of it. The assessor may decide to concentrate on either the investigations the trainee requested or the ethical concerns the patient highlighted. The assessor is interested in assessing the understanding the reasoning skill of the trainee.

### **Mini Clinical Evaluation Exercise (Mini-CEX)**

Due to therapeutic responsibilities, teaching and learning are frequently unstructured in clinical settings. Therefore, technologies that make the processes clear for both students and teachers are crucial for learning. The usefulness of Mini-CEX in identifying areas of strength and weakness across a range of topics demonstrates its significance in finding teaching opportunities<sup>[1]</sup>. Mini-CEX records a moment of a doctor-patient encounter. It is intended for an assessor to give trainees comments on abilities crucial to delivering high-quality clinical care while watching a real-world clinical interaction. Usually, this assessment only looks at one aspect of the clinical contact, such obtaining a history or doing a clinical examination, and it takes place in a clinic or ward. The evaluation is documented using a standard proforma, and the intended standard is that of a trainee at the completion of the relevant training phase<sup>[9]</sup>. Mini-CEX is led by trainees. It is important to inform patients that this type of exercise is being performed. After every mini-CEX interaction, strengths, areas for improvement, and agreed-upon measures should be documented.

The encounters are intended to take about 15 min and the trainees are evaluated several times by different faculty members. Nine-point rating scale is used where 1-3 is unsatisfactory, 4-6 is satisfactory and more than 7 is expert. The interviewing skills, physical examination, professionalism, clinical judgement, counselling, organisation, and efficiency are few mini-CEX encounters.

The mini-CEX is comparable to a test conducted in a classroom for a medical setting. Its purpose is to identify trainees who are performing poorly and to document their shortcomings. This documentation serves as justification for the decision made by the

educator regarding the trainee. The tool was more helpful to trainees who had received instruction in using Mini-CEX in medical school than it was to students who had not <sup>[10]</sup>.

### **Evaluation of Clinical Events (ECE)**

One new tool is ECE. It offers an approach to evaluate how well the trainee performs their responsibilities in challenging tasks, which frequently call for collaboration or communication with other qualified employees. Examples include presenting a case at a multidisciplinary team conference, evaluating and reporting diagnostic material clinicopathologically, and participating in quality assurance and audit procedures in both clinical and laboratory settings<sup>[75]</sup>.

### **Multi-source Feedback (MSF)**

Sometime referred to as 360° feedback, MSF is a technique for getting organized feedback from trainee-related staff members who have the chance to observe them in action. These employees may include their managers as well as peers and those the trainee oversees or even looks after in terms of their relationships and conduct in a professional manner.

Anonymized feedback is provided with a list of qualities or behavioural characteristics for the learner to consider and make the necessary corrections.

Additionally, the trainee contributes their own review of their own performance. It offers fair criticism on the trainee's conduct and proficiency in clinical settings that the supervisor might not be able to see first-hand. Mini peer assessment tool, team assessment of behaviour, and patient satisfaction questionnaire are the other tools used to assess this domain of assessment.

### **Mini-peer Assessment Tool (mPAT)**

- mPAT encompasses the integration of ideas about a trainee's performance in a range of competence domains from their colleagues<sup>[76]</sup>. This assessment strategy gathers confidential feedback from eight peers evaluating 16 aspects from the aforementioned fields
- diagnosis and effective use of the investigative tools at hand
- time management
- Stress management and work-life balance
- Successful communication
- recognising one's own limitations

### **Team Assessment of Behaviours (TAB)**

TAB is a type of multisource feedback evaluation used in the UK Foundation Curriculum for medical trainees <sup>[11]</sup>. Based on the GMC's guidelines for professional conduct, TAB has established the following four domains:

- Building and preserving relationships and a professional rapport with the patients
- Effective verbal communication; teamwork and team leadership; and guaranteeing availability and accessibility.

To assist individuals in improving their performance, TAB is utilized as a formative and summative tool. A minimum of ten results are required for this assessment tool to be considered legitimate and trustworthy. Since ratings differ greatly by staff group, a suggested mixture of raters is offered.

### **Patient Satisfaction Questionnaire (PSQ)**

As part of an appraisal process, PSQ can offer formative comments on a physician's professional performance <sup>[12]</sup>. Patients' opinions are gathered through a standardized questionnaire. At least once every five years, doctors are supposed to receive feedback, consider it, and, if necessary, use it to guide their future professional development. The demographics of the patients and the means used to administer the questionnaire (such as postal, telephone, or proxy responses) may have an impact on the final ratings when patients evaluate doctors or bigger health care systems. When peers evaluate the work of other doctors, the entire evaluation process may be compromised by the rater's personal opinion, the length and type of the rater-examinee relationship, and the rater's knowledge with the doctor's practice<sup>[13]</sup>.

According to numerous publications in the body of current literature, multisource feedback can objectively evaluate critical qualities such as professional expertise, collegiality, interpersonal skills, communications skills, and the capacity to advance in the medical sector. However, multisource feedback has pitfalls of its own. Several studies have shown that responses tend to be skewed towards positive assessments of doctor performance by the patients and some studies have shown dissatisfaction about the ability of multisource feedback, patient feedback in particular, in identifying the underperforming doctors.

### **Portfolio**

It keeps track of every record produced by workplace assessments, clinical experiences, reflections, meetings attended that were pertinent, informal or formal training activities, important situations, etc. This tool's structure and goal are to demonstrate the total amount and quality of learning and advancement. Paper-based (diary, etc.) or digital (e-portfolio, etc.) portfolios are both acceptable. These can be categorised as reflective, developmental, assessment, or showcase portfolios depending on their functioning.

**Showcase Portfolio:** These portfolios, which emphasis the portfolio as a product, are also sometimes referred to as formal portfolios, professional portfolios, or career portfolios. After learning occurs, the content for showcase portfolios is prepared, frequently with student reflection. "Collect, Select, Reflect, Connect" is a motto that certain schools embrace (PDF Hughes, 2008). The connect component is an intriguing one because it entails sharing student work with others (perhaps outside of the teacher) and actively seeking feedback. The best accomplishments or learning proof from a student are frequently shared in the showcase portfolio. In most cases, students get to choose what gets published.

**Assessment Portfolios:** The assessment portfolio is used to record a student's learning or show that they have mastered certain curricular components. Reflective remarks will now concentrate on how artefacts match learning goals. In comparison to a showcase or process

portfolio, these sorts of portfolios could be more formal. An assessment portfolio may be very helpful for teachers and administrators to see proof of learning inside the classroom, but it may not be as helpful for students' overall growth. Commonly, certification programme or even requirements for receiving a degree include assessment portfolios.

**Reflective Portfolio:** reflective portfolio serves as a summary of the knowledge and experiences a student has acquired through practical tasks. It is intended to evaluate the student's involvement in their fieldwork and their aptitude for applying theory in practical situations. It can include work samples, critical events, student's own perspectives, evidence of achievements and journal entries.

**Development Portfolio:** Students choose their works and record the works that demonstrate development or change of learning over time. The learning process is highlighted in the development portfolio.

### **Entrustment Aligned Pathology Assessment Instrument for Intraoperative Consultations (EPA-IC)**

EPA-IC is developed in 2015 and introduced at Western University's Anatomical Pathology training program in 2016 [14]. It was used by clinical examiners as part of the routine formative WBA of PGY-2 to PGY-5 residents' performance of intra-operative consultations. The EPA-IC is an 11-item assessment tool that evaluates residents' abilities to perform intraoperative consultations from case preparation through the post-procedure plan. Aspects of patient safety, such as tissue handover, communication, and teamwork skills, were also taken into consideration in addition to diagnostic interpretation and technical competence. Eight things were graded on a 5-point scale, one was a yes/no question about whether the student was prepared to practise independently, and the other two were open-ended inquiries about one particular component of the case that went well and one that needed work. The rating anchors, which varied from 1 = "I had to do" (i.e., the trainee needed full hands-on guidance or did not complete the procedure) to 5 = "I did not need to be there" (i.e., the trainee had complete independence and is practice-ready), were based on the rater's assessment of the trainee's required supervision and support level<sup>[14]</sup>.

### **Workplace Based Assessment in Emergency Department**

Emergency medicine (EM) educators and programme leadership face significant hurdles because of the special practise environment of the ED. During the 2012 Academic Emergency Medicine (AEM) Consensus Conference on Education Research, a breakout session on assessment of observable learner performance in EM covered a number of these issues, including the viability of conducting direct observations while providing patient care and supervising acutely ill patients. The essay that resulted from this breakout session described a number of methods for evaluating learner performance, including both direct and indirect methods (e.g., resident portfolios, procedure logs, self-reflection) <sup>[15]</sup>. Direct observation tools used in ED includes ACGME EM Milestones, Observed Structured Clinical Exercises (OSCE), McMaster Modular Assessment Program (McMAP), Queen's Simulation Assessment Test (QSAT), and the mini-Clinical Evaluation Exercise (mini-CEX). The Ottawa Emergency Department Shift Observation Tool, the Reporter/Interpreter/Manager/Educator (RIME) framework, the Standardized Direct

Observation Tool (SDOT), the Critical Care Direct Observation Tool (CDOT), the Mini card, a non-milestone-based end-of-shift evaluation, checklists, a global rating scale, and the Resuscitation Assessment Tool (RAT) were among the other tools.

### **ACGME EM Milestones**

Each speciality designed the Milestones, a framework for evaluating resident development, to address the six basic skills established by the ACGME. It assesses the diverse array of technical and non-technical skills. It is the most commonly used tool. It is utilized by all Emergency Medicine residency programs as part of their assessment of residents and is required to be reported to the ACGME for reaccreditation <sup>[15]</sup>. Many writers cautioned against developing end-of-shift or simulation assessment systems utilising ACGME milestones. According to Dehon et al., there was little correlation between end-of-shift milestone scores and clinical competency committee evaluations at one site and all resident levels achieved level 3 milestones at around the same rates. Alternatively, Dayal et al. discovered that milestone scores rose 0.52 levels year. According to Lefebvre et al., the learner assessment scores given by the clinical competency committee increased when narrative comments were included to milestone scores on end-of-shift tools <sup>[15]</sup>.

### **Critical care Direct Observation Tool (CDOT)**

It is focussed on critical care interventions at clinical setting<sup>[15]</sup>. Mapped to milestones. Includes a qualitative comments box. But it is limited to yes, no, or Not Applicable responses. And the CDOT has poor inter-rater reliability<sup>[15]</sup>.

### **Checklists**

Both clinical settings and simulations make use of checklists. Every clinical presentation is comprised of a checklist. and might provide a space for qualitative comments. It can also be used to assess ACGME milestones if it is matched to them. Its inter-rater reliability is good. For every primary complaint, a unique checklist must be developed. focused primarily on particular care features or presentations. Frequently, there are only three possible answers: yes, no, or unclear <sup>[15]</sup>. Qualitative remarks differ depending on the checklist.

**Global Breaking Bad News Assessment Scale** Short and easy to complete. Study tool can be modified to include a qualitative comments box. Only assesses delivery of bad news. Responses limited to yes or no. but resident skill increases after each encounter <sup>[15]</sup>.

**Global Rating Scale:** Fewer questions. Faster to perform. Can be combined with other direct observation tool. It has good inter-rater reliability. For clinical judgement and communication. It relies heavily on gestalt and it has less granular assessment of components and it has no qualitative comments.

**Local End of Shift Evaluation:** It can includes assessment of technical skills and some non-technical skills (e.g., professionalism, interpersonal skills). The Categorizations are general with limited specific examples. Not all tools have qualitative comments <sup>[15]</sup>.

**McMAP (Mc Master Modular Assessment program):** The tool is learner-centered <sup>[15]</sup>. The ACGME and CanMEDS Frameworks have been linked to individual clinical examinations.

The tool includes required written comments and behavior-anchored scales. Because there are 76 distinct tests in the tool, there can be a larger learning curve. Depending on the patients seen, certain components might not be observable. Students might steer clear of difficult assignments or ones in which they lack proficiency. Some elements that are more difficult to assess may be avoided by faculty.

**Minicard:** It follows with an action plan and remarks for every single assessment item. Results may be skewed if trainee level is included in scoring descriptors.

### **O-EDShOT** (Ottawa ED Shift Observation Tool)

It was created especially for the ED environment using input from residents and instructors. Although there is some validity data to support its usage, more research is required. The Ottawa Emergency Department Shift Observation instrument was introduced as an entrustment-based instrument to assess a resident's capacity to manage the ED. includes a space for qualitative comments about one's strengths and shortcomings. Regardless of the treatment area (high, medium, or low acuity), it can be applied<sup>[15]</sup>. The O-EDShOT was created to assess a set of skills over the duration of the whole shift, not simply one patient encounter. O-EDShOT's ability to distinguish between residents with different levels of training and the fact that scores did not differ according to the ED treatment area show that the test can be used to evaluate a resident's capacity to manage an ED shift regardless of the acuity area to which they are assigned <sup>[15,16]</sup>. Furthermore, the O-EDShOT is practical and useful for promoting feedback aimed at progressing toward independent practise, according to front-line teachers and residents. The normative scale that was previously in use required the assessor to transform judgements of the trainee's performance into abstract anchors based on concepts that were poorly comprehended for the resident's level of training (e.g., below, meets, surpasses expectation), but faculty and residents reported that the O-SCORE was more pragmatic and objective than that scale <sup>[16]</sup>. The O-EDShOT could be used to evaluate specific entrustable professional activities in the transition to practise stage of training that relate to managing the ED and help the clinical competency committee make decisions about a resident's readiness for independent practise because it captures the essential skills of an EM physician. Literature implies that entrustability measures are construct-aligned and reflect the goals of the clinician-assessor, supporting these ideas. Additionally, The language of entrustment anchoring has been organised around movement towards autonomous practising, making the connection between clinical assessment and improving skills in the workplace clearer, trainees perceive ratings on entrustability scales to be more transparent and justified <sup>[16]</sup>.

**QSAT:** Provides a framework that can be customized to each specific case. Each QSAT would need to be individually designed for each presentation. Studies limited to the simulation environment.

**RAT:** Builds upon QSAT with Entrustable professional activities targeted towards resuscitation management. Designed using a modified Delphi study with experts. RAT was positively correlated with entrustment scores. Only assesses resuscitation management.

**RIME:** Easy to use. Can be combined with other tools. There is positive correlation between RIME category and clinical evaluation scores.

**SDOT:** Both technical and certain nontechnical skills (such as professionalism and interpersonal skills) are assessed. In certain patient visits, a number of components could not be appropriate. does not offer a qualitative commenting option. reduced accuracy in comparison to alternative tools. might take longer than other methods of direct observation<sup>[15]</sup>.

### Daily Encounter Cards

Daily encounter cards (DECs) have developed into a key type of WBA utilised in the ED setting to evaluate trainee performance. DECs make it easier to evaluate several critical abilities at once using performance data gathered over the course of a shift. The supervisor often completes DECs at the conclusion of each shift, eliminating recall bias while also providing a catalyst for frequent formative input and for repeated assessments of performance over time. Evidence reveals that, despite their widespread use, DECs' assessment documentation quality is subpar<sup>[77]</sup>. According to a research by Bandiera and Lendrum<sup>[17]</sup>, DECs were vulnerable to leniency or range limitation effects, in which supervisors gave "inflated" or excessively positive evaluations that resulted in indiscriminate ratings. According to Sherbino et al., front-line supervisors' poor comprehension of the DEC items led to low reliability and shaky validity of the results<sup>[18]</sup>.

O-Score

An evaluation tool for WBA, the Ottawa Surgical Competency Operating Room Evaluation (O-SCORE), focuses its grading scale on a specific set of entrustment anchors. This score has good psychometric properties<sup>[16]</sup>.

### International Experience

I The Chinese government introduced the resident standardisation training (RST) programme in 2015<sup>[30]</sup> in response to CBME and to guarantee high competence levels among practitioners for high-quality healthcare. The mini-CEX is advised as a formative evaluation instrument in the Chinese RST programme due to its exceptional dependability, efficacy, simplicity, and multifunctionality. The mini-CEX has been utilised for the evaluation of interns, residents, and postgraduates in clinical medicine and medically related professions during the three-year RST programme after being specially modified based on the original format<sup>[31]</sup>. The Mini-CEX was developed, piloted, and evaluated in the USA and is now widely used to assess doctors on American Residency programmes<sup>[32,33]</sup>.

Mitchell. C et al utilised anonymized records for 1646 trainees in a single UK postgraduate deanery to conduct a retrospective observational analysis. The e-portfolio database's data for WPBAs conducted between August 2005 and April 2009 were taken. All of the results from the mini-clinical evaluation exercise (mini-CEX), case-based discussion (CbD), direct observation of procedural skills (DOPS), and mini-peer assessment tool (mini-PAT) evaluations completed by trainees in FP years 1 and 2 were included in these data. and discovered that 92 of the 1646 trainees had been classified as having a problem. There was a correlation between identified training challenges and lower mean CbD and mini-CEX scores for trainees who experienced difficulties<sup>[34]</sup>.



In order to assess their experiences with foundation programme WPBAs, F2 doctors in Northern Ireland took part in an electronic survey created by McKavanagh P et al. Electronically presented survey questions were used, and Turning Point technology was used to compile the replies. F2 doctors seek more chances for valuable consultant interaction with prompt feedback, and they discovered that the WPBAs provided the foundation doctors with a one-on-one learning opportunity with their supervising consultants. Since the current WPBA procedure lacks integrity, they recommended improving its implementation <sup>[35]</sup>. McLeod *et al.* <sup>[36]</sup> concluded that, DOPS provides an overall insightful perspective of a students' procedural skills assessment in undergraduate medical students in UK. Morris *et al.* <sup>[37]</sup> concluded that DOPS possesses significant positive feedback, results in improving competence-based learning among interns.

An example of DOPS is in use in UK foundation programme. They have a list of procedures that is being done routinely which comprises intubation, the placement of a nasogastric tube, venepuncture, arterial blood sample, and an electrocardiogram. Students are evaluated on their skill, understanding of indications, asepsis, and communication, and other aspects of the particular procedural skills. The trainee must document how often they are being assessed. Of late, the ratings were eliminated, and the learner now receives free text evaluation and comments rather.

A variety of workplace-based assessments (WBAs) were implemented into Chinese postgraduate medical education after the implementation of Standardized Resident Training Program in 2017. In a study conducted at a teaching hospital at southeast China, perceptions of the mini-CEX among Chinese trainees and supervisors, and their understanding and attitude towards this assessment tool was studied using the first-year postgraduates. It was found that the mini-CEX has positive educational value when appropriately administered in clinical training, although results may be limited by learner and supervisor attitudes and knowledge. In that study, almost every interviewee agreed that feedback was generated via observations of trainees, which were made regularly. While trainers thought that being observed contributed to a sense of learning, supervisors used observation to pinpoint particular areas to improve. Because the mini-CEX had such a direct impact on their training, trainees regarded it more highly than supervisors did. Even if some researchers have talked about low involvement with the mini-CEX, most students think their peers were still being motivated <sup>[1]</sup>.

Medical graduates in the UK enter the Foundation Programme that constitutes the first 2 years of postgraduate training. Each year, it is expected that trainees undertake a defined number of DOPS, mini-CEX and CBD assessments <sup>[38]</sup>.

Andrea C. Lorwald et al. in their systemic review found that Four themes (context, users, implementation, and outcome) and nine subthemes (time for Mini-CEX/DOPS, usability of the tools, supervisors' knowledge of how to use Mini-CEX/DOPS, supervisors' attitude towards Mini-CEX/DOPS, trainees' knowledge about Mini-CEX/DOPS, trainees' perception of Mini-CEX/DOPS, observation, feedback, and trainees' appraisal of feedback) were identified as influencing factors on the educational impact of Mini-CEX/DOPS <sup>[6]</sup>.

A hermeneutic review by Shaun Prentice et al from the articles limited to Australia, New Zealand, UK, Canada, the Netherlands, and Scandinavian countries to maximise the

comparability of results, as each of these countries have similar models of postgraduate medical education and typically recognise each other's qualifications. The review showed that assessing in the context of the job was also believed to encourage learning and improve the application of knowledge. Another topic related to how WBAs encouraged and even mandated the provision of feedback in order to fill a training deficit in medicine. Another benefit is that WBAs make use of the assessor's expertise, in contrast to other assessment forms. WBAs enable the early detection of struggling trainees, lowering the possibility that patients may receive subpar care and assuring the effective and prompt allocation of remediation resources. It also found that the use of WBAs in heavy workloads is hampered by time constraints that are prohibitive, especially for some technology platforms and the fact that many WBAs are needed to make trustworthy judgments <sup>[39]</sup>.

According to Vasiliki Andreou et al., in a three-round web-based Delphi survey, a group of experts) were asked to evaluate the feasibility of the CanMEDS (Canadian Medical Education Directives for Specialists) key competencies for workplace-based evaluation on a 5-point Likert scale <sup>[40]</sup>. The Royal College of Physicians and Surgeons of Canada developed the CanMEDS competency framework initially for setting educational outcomes for graduate medical education, and it is today the most extensively used and approved framework within medical curricula worldwide <sup>[41, 42]</sup>. Different outcomes are identified and described as competences that physicians should develop to practise patient-centred care in the CanMEDS framework. The seven positions of medical expert, communicator, collaborator, leader, health advocate, scholar, and professional are thematically used to organise these talents <sup>[41]</sup>. In accordance with the framework, competencies are divided into two levels: level one contains the key competencies, and level two the enabling capabilities. Curricular adjustments are required to facilitate the application of CBME in postgraduate medical training. CBME mandates, among other curriculum modifications, the establishment of learning outcomes that promote educational continuity as well as the alignment of learning goals with learning and assessment activities <sup>[43]</sup>. According to Vasiliki Andreou et al., the panel was asked to rate 12 CanMEDS key competencies for feasibility and 15 for consistency of assessment in the workplace and the panel reached consensus for 6 out of 12 CanMEDS key competencies for feasibility and for 4 out of 15 for consistency of assessment <sup>[40]</sup>. Although an outcome-based strategy does not necessarily result in learning, it unquestionably creates all the conditions that do. The prerequisites for implementing competency frameworks include clearly defining the learning outcomes that should be attained by learners, giving them opportunities to practise these outcomes in a variety of settings, creating opportunities for assessment and feedback, and fostering reflection on personal performance. According to Vasiliki Andreou et al., <sup>[40]</sup> that not all CanMEDS key abilities could be clearly matched to observable behaviour, which has implications for the assessment feasibility. In comparison to competencies under other CanMEDS roles, such as "Medical Expert," "Communicator," and "Scholar," some important CanMEDS competencies under the "Leader," "Health Advocate," and "Professional" roles received noticeably lower ratings (50%). Large variations in rating scores may indicate that the panel had trouble connecting how those CanMEDS competencies may be applied to and translated into workplace assessment tasks. Lack of consistent and concrete descriptions in undergraduate medical education has been blamed for the difficulty in implementing the CanMEDS essential competencies in workplace-based evaluation. Implementing the Can- MEDS non-medical competences have been challenging in postgraduate medical education, which has been linked to lack of training for workplace-based assessors. The study also found that there is a discrepancy between the CanMEDS

competency framework's original goal and its usefulness for workplace-based evaluation. For organising workplace-based assessment and capturing medical competence, the CanMEDS essential competencies may provide a good place to start <sup>[40]</sup>. The framework still has to be further developed and contextualised in order to support observations of trainee behaviour involving all seven CanMEDS roles during clinical practise. To assemble more data on the CanMEDS framework, future research should examine implementation issues in various healthcare contexts and settings <sup>[40]</sup>.

Barret et al did a qualitative study to find out the perception of the trainer and trainee about the WPBAs <sup>[29]</sup>. Teunissen's "experience, trajectories, and reifications" serve as the study's conceptual underpinning (ETR). The framework's goal is to explain how an individual's varied encounters with a given learning concept cause them to follow a specific, unique learning trajectory that eventually helps them learn. The study illustrated WPBA as a mere "tick-box" activity where WBAs were finished at the end of a year or training position to satisfy requirements. Forms were rarely completed, and WBAs were not set up prospectively, following a case-based discussion or during patient-trainee interactions. The opinions of the WBAs' sincerity were similarly impacted by these recurring tendencies. A mini-CEX was perceived by trainees as a "set-up" rather than a live observation that just involved taking a patient's history or doing a physical examination. In the study, Most trainers thought trainees should come to them to finish WBAs and remembered having to "chase" trainees to finish them. Additionally, only the Trainers connected WBA experiences to the e-portfolio technology, and their evaluations of the learning value were more closely related to time commitments and inadequate technological configurations. One instructor felt there wasn't enough time to assess communication issues even though it was their obligation to ensure technical skill mastery <sup>[29]</sup>. Nesbitt published the views of University College London students in the UK. With 31% of the cohort indicating that WBAs interfered with their teaching time with the evaluating doctors, there was disagreement within the cohort as to whether WBAs were effective tools to promote contact with seniors <sup>[44]</sup>. Al-Kadri described the experience of a Saudi Arabian university where WBAs are now required. According to the authors' summary, their students believe that the validity of their WBA results depends on who is evaluating them <sup>[45]</sup>. As per Ali, although participating in WBAs did not encourage students to have more favourable opinions of WBAs, those who understood its purpose did show an appreciation for WBAs' contribution to their education. This discovery is consistent with the finding that trainees' inadequate comprehension of WBA tools is a major issue. It was interesting to note that people with experience weren't any more likely than those without experience to show a positive comprehension, indicating that knowledge of WBAs may be acquired without experience. Although having experience with WBAs is crucial, comprehension also seems to be linked to students' creation of favourable impressions, and experience alone may even be detrimental. The money spent by medical colleges could help change how people view and interact with WBAs at work <sup>[38]</sup>.

### Indian Scenario

A few institutions in our country are using WPBA tools like the mini-CEX, DOPS, or tools that are similar to them. They are used independently rather than as a part of a deliberate WPBA programme, though. Initial reports on practicality and teacher and student acceptance are encouraging. Kamat C. et al conducted one-year prospective interventional study in the department of anaesthesiology, where in 55 postgraduate students and 21 faculty from

department of anaesthesiology participated in the study. After orientation of faculty and postgraduates, the study was conducted to study the Direct observation of procedural skills (DOPS) effectiveness as a tool for anaesthesia assessment. Three commonly performed core skills of anaesthesiology, i.e., Spinal Anaesthesia, Epidural Anaesthesia, and Endotracheal intubation were the selected procedural skills for the assessment and found that there was a significant improvement in the post-DOPS scores indicating the improvement in the procedural skills <sup>[46]</sup>.

Hill and colleagues used DOPS assessment tool during ultrasound-guided central line insertion in anaesthesia department and developed a DOPS tool focusing on key components of safe practice of central line insertion <sup>[47]</sup>.

Kumar *et al* conducted a study among OBG postgraduates and concluded that DOPS offers high level of satisfaction and improvement in surgical skills for OBG postgraduate trainees. The advantages of DOPS noted were, provision of rapid and constructive feedback in the form of both marks and comments <sup>[48]</sup>.

John Roger Barton *et al.* evaluated DOPS assessment on practitioners in endoscopy colonoscopy and suggested that use of DOPS can be expanded for relicensing or recredentialing and should be considered for assessment of competence in all clinical areas <sup>[49]</sup>.

Liaqat Ali *et al.* did a comparative study among urology postgraduates while performing various urology procedures like TURP, Cystoscopy, URS, etc., and demonstrated that DOPS is effective assessment tool and improves the skills of urology postgraduate residents <sup>[50]</sup>.

Profanter and Perathoner in their benchmark study of prospective randomized trial in small groups of undergraduates, concluded that DOPS is an efficient tool in teaching clinical skills as compared to OSCE <sup>[51]</sup>.

Joshi MK *et al* studied about the acceptability and possibility of using a mini-clinical evaluation activity as a formative evaluation tool for workplace-based evaluation of surgical postgraduate students. Over the course of seven months, nine faculty members conducted a total of 60 mini-CEXs with 16 second-year postgraduate students from the department. During this time, each resident had a minimum of three and a maximum of five contacts. Interviewing skills, physical examination skills, counselling skills, clinical judgement, and an overall impression were all tested throughout the mini-CEX. They discovered that while most postgraduates enjoyed the laid-back atmosphere during the exercises, a few of them became anxious when faculty members were watching them while they interviewed or examined the patient. They also discovered that all postgraduates agreed that the mini-CEX could be used for internal assessment <sup>[52]</sup>.

Khalil S *et al* conducted a cross-sectional study from August 2015 to January 2016 in the paediatric department of a teaching hospital in India to study the adoption of a Mini-Clinical Evaluation Exercise (Mini-CEX) Programme to Evaluate Postgraduate Paediatric Trainees' Clinical Competence. A total of 20 final year postgraduate students were assessed with a total of 112 Mini-CEX encounters conducted by six faculty members. Ninety percent of the participating students felt that Mini-CEX changed their attitude towards teaching, and it

should be included as a routine in postgraduate teaching. Only 25% thought that it induced anxiety in them. On assessment of faculty perception of Mini CEX, all thought they had a valuable experience and the teacher's feedback would improve students' performance, whereas 50% were doubted whether it was a valid method of assessment <sup>[53]</sup>.

Gupta et al studied about the Mini-clinical Evaluation Exercise Acceptability and Feasibility as a Learning Tool for Paediatric Postgraduate Students. The study was conducted from May 2016 to October 2016 in the department of paediatrics at a tertiary level medical university in Northern India. There were a total of 87 mini-CEX contacts with 29 residents, 3 SR, and 13 staff members. Residents who were asked open-ended questions said that the mini-CEX increased their clinical abilities, promoted their personal growth, and allowed for better one-on-one student-teacher interactions. They asked for more of such encounters even though some of them thought it was hectic in between their work schedule. Faculty perceived that they found it useful for improved learning for themselves also. However, the faculty also felt that it requires more time to assess the students so shall not be possible for a larger batch. Some faculty concerned about the subjective bias. However, instructors and residents both suggested integrating Mini-CEX in the curriculum <sup>[54]</sup>.

Sethi et al conducted a study to find the effectiveness of mini-CEX as a formative assessment tool in the postgraduates of Psychiatry from April 2019 to September 2019 at the Pt. BD Sharma PGIMS, Rohtak, Department of Psychiatry. The trainee will rotate between all five assessors (including the investigator) according to a tentative schedule of mini-CEX sessions, which also ensures that no two sessions will be with the same assessor back to back. A minimum of two weeks has to pass between two mini-CEX interactions. After the study was finished, feedback from professors and graduate students was obtained using feedback forms specifically created for the purpose, and these forms were then verified by the senior faculty members. Postgraduates were asked to complete Learning Self-Efficacy Scale also for clinical skills. Comparing the mini-CEX to other teaching techniques, 61% of postgraduate students said it was a superior method, 36% said it enhanced their clinical abilities, and 64% said it increased their confidence. A third of them claimed to have felt stressed and anxious throughout the interactions. A large proportion (88%) of residents felt that the mini-CEX should be regularly employed in academic contexts. Similar to the previous point, the majority of consultants who had already received training in a variety of assessment exercises discovered that mini-CEX was superior to more conventional assessment techniques like bedside case discussions and outpatient departments because it offered quick feedback and had a structured format for evaluation. The majority of the faculty members (70%) suggested mini-CEX as a suitable formative evaluation tool and that it should be included early in the training, despite the drawbacks of subjective bias and time restrictions. During comments, the majority of the students expressed their agreement that this activity had improved their ability to learn and their clinical skills. The faculty stated that mini-CEX was superior because it offered quick response and had a focused, organised pattern of evaluation <sup>[55]</sup>.

Batra et al conducted a study on the method for evaluating residents in the department of surgery is the "Mini Clinical Evaluation Exercise" (Mini-CEX). From May 2019 to September 2019, the study was carried out in a tertiary care facility where the department of surgery had residents. In this study data gathering, diagnosis, therapy, and counselling were the focus areas. There was significant improvement in interviewing skills, physical

examination skills, professionalism, and counselling skills after mini-CEX but no difference was obtained in clinical judgement and organisation efficiency skills <sup>[56]</sup>.

### **Advantages of WPBA**

1. The assessment of the trainee in the workplace reduces the artificiality of traditional method of assessment.
2. The workplace-based assessment maintains the content and context specificity of the assessment.
3. The workplace can be used as a sampling situation where the trainee will work after qualifying.
4. It is feasible to assess the competency holistically.
5. A variety of skills can be assessed across a range of contexts and scenarios.
6. Assessment can be incorporated into the regular activities without detracting from it.
7. It assesses the 'does' level of Miller's pyramid.
8. Assessment of professionalism is possible.
9. Assessment is for learning rather than assessment of learning.
10. It emphasizes the learning of communication with patients, peers, communication with paramedical workers and team management than focusing only on treating the illness.
11. Many assessors can evaluate and provide inputs for a single trainee.
12. It is learner centric by providing the opportunity for the learners to actively participate in patient selection and decision making.
13. WPBA modulates the learning process by providing feedback for the learner to improve.
14. It is a longitudinal assessment rather than a midterm or end of posting assessment.
15. The feedback is more effective when it is given in relation to a particular task. It promotes careful observation and feedback during work, which can facilitate problem- and context-specific learning.
16. Since WPBA assessments are done using multiple assessors at multiple times, it is very much reliable.
17. WPBA had a positive educational impact (Kirkpatrick level 1) and could lead to modifications in attitudes (Kirkpatrick level 2a) or even changes in behavior (Kirkpatrick level 3) and helped to improve clinical skills (Kirkpatrick level 2b).

Although there were several factors that affected usefulness, WBAs were generally thought to be helpful because they offered learning opportunities, knowledge enhancement, curriculum coverage, formalisation of training, reflection and feedback tools, and adjuncts to educational supervisor (ES) and clinical supervisor (CS) reports. The style of validation and the time of validation, as well as the engagement of trainee and trainer and their alignment with one another, appeared to be key factors in the usefulness of the WBA <sup>[57]</sup>.

### **Limitations of WPBA**

1. WPBA are supplement to formative assessment rather than replacement to conventional methods.
2. Student who performed well in initial interactions may become overconfident while the weaker trainee can be deterred by the first few interactions and refrain from asking for feedback.

3. Since senior and experienced assessors will give lower but more accurate ratings, the trainee may prefer less experienced assessors to get more ratings.
4. The reliability of the WPBA tool depends on how the tool being used and it requires faculty training for the effective use of these tools. The largest hurdle in implementing WPBA is faculty training. Clarity regarding what to assess and what standards to anticipate, as well as the skill of providing useful comments, are two key areas where assessor training is crucial.
5. Training in assessment clarity and norms will lessen the likelihood that assessors would overlook subpar performance or a crucial ability.
6. Any trainer must have training in providing constructive criticism because feedback greatly aids in learning. If the feedback is not given in an acceptable, constructive manner with ideas for development, the benefit of the entire process may be lost.
7. The dual responsibilities of faculty members as teachers and assessors may conflict. This could lead to reluctance to record negative or mediocre evaluations.
8. To make these tools more acceptable, trainees must also be sensitised and given the positive consequences of feedback.
9. There may be inertia in introducing new assessment tools like WPBA.

There is an inherent bias in the system because many people only give assessments that describe what went well, leaving out the ones that did not. This results in a lack of documentation of all learning experiences, which reduces the content validity of the instrument.

Lack of time to validate, delayed scheduling of validation, e-mail validation rather than face-to-face validation, varying quality of input, lack of follow-up on feedback, tick-box exercise, importance of quantity over quality, and loss of accuracy are some of the challenges identified.

Researches revealed that there was a lack of time for both the trainee and the trainer, difficulty in finding an assessor who was willing and qualified, a lack of enthusiasm, a lack of trainer training and knowledge regarding the requirements of the trainee, an emphasis on quantity rather than quality, uncertainty about validity, a delay between the event and the completion of feedback, and a lack of understanding of the purpose of WBAs are the barriers in using WPBAs as assessment tools<sup>[57]</sup>. Research shows that learners may not know how to use the input from WBAs, while assessors may not know how to conduct WBAs. Stakeholders also claimed that the lengthy implementation times for WBAs made it difficult to integrate them into existing workflows. Stakeholders questioned the evidence supporting WBAs, particularly their content and discriminant validity, and claimed to find the frameworks, definitions, and tools for WBAs to be complex, ambiguous, divisive, and insignificant. For trainees, who were worried about being evaluated under unnatural circumstances, which were frequently brought on by the stress of observation, the tension between summative and formative assessment was obvious. Unsettlingly, this conflict seems to jeopardise formative WBAs as well, with trainees believing that "formative" WBAs are being employed for summative evaluations. The assessor-trainee relationship, the assessor's assessment literacy, and the competency of the trainees all have an impact on how frequently trainees choose assessors who would favourably evaluate them. The preferences and attitudes of stakeholders (particularly assessors) regarding particular WBAs have an impact on both the choice of a WBA and subsequent engagement.

Assessors with inadequate clinical expertise may consider trainees with equally weak clinical expertise as competent if they use their own level of clinical expertise as the benchmark by which they rate a trainee's competency. Additionally, prejudices held by assessors may influence how they remember observed behaviour and make judgments about trainees. Similar to this, those who find evaluations uncomfortable and/or who lack the training necessary to support their choice may "fail to fail" trainees who require additional support. The assessor's position within the evaluation environment also affects the decision; comments will be influenced by a person's profession (for example, nurse vs. clinician), seniority, and perceived role as an assessor or instructor [39].

Through workplace-based evaluation, the supervisors have the chance to analyze the trainee's clinical practice. However, many trainees view workplace-based assessments as a test to pass rather than an opportunity to learn, and they may adjust their practice expressly for these exams. Lack of confidence in their supervisor may be one cause of this inauthenticity. Due to the advent of entrustment as a recommended method for supervisors to assess trainees' progress toward competent practice, most of the research and criticism have concentrated on supervisors' need to trust trainees. The trainers trust the trainees who are believed to be capable, diligent, honest, and responsible. They look for confidence, perception, and a willingness to solicit help, take criticism, and willing to learn from feedback. Domain et al discussed how workplace-based assessment affects how much trainees trust their supervisors. Instead of purposefully assessing trust moment by moment, trainees made trust decisions naturally. Although they acknowledged the significance of this feeling of trust in their training and assessment, participants needed encouragement to explain how they came to trust their supervisor. This may be due to the unconscious nature of these decisions. The limited trust that trainees first displayed—primarily based on the supervisors' perceived duty to perform their duties—was followed by an increase in trust that was influenced by experience. In their research, Domian et al. discovered that Bourdieu's idea of "feel for the game" perfectly captured how trainees trust decisions in workplace-based assessments were made intuitively in reaction to supervisor behavior [58,59]. These fast-paced interactions give little room for thought, but practice gives players a sense of the game that helps them maneuver through these situations. Our findings imply that trainees may experience emotional costs as a result of the trial-and-error process needed in learning to navigate this complexity. We believe it might add to the emotional strain that impression management is said to involve [60]. While increasing trainees' knowledge of the learning process and its emotional effects may be beneficial, this learning may be essential to trainee development. The study also found that Positive effects result from supervisor credibility building trust among trainees. Impression management is regarded by students as essential in surgical training in order to access chances for practice and learning; yet, it adds emotional work, can impede learning, and can jeopardize patient care. Trust made it possible to communicate feelings honestly [58].

Medical students seem to have unfavorable opinions on WBAs. Although having experience with WBAs had little bearing on their viewpoint, having knowledge of WBAs appeared to have an impact on attitudes. This shows that how medical students are exposed to WBAs should be carefully considered in order to guarantee that it supports the development of passion and positivism that the students can carry into their professional lives [38].



The misalignment between how supervisors cognitively form evaluations of the trainee and how they are asked to document these judgments poses a serious danger to the validity of many extant WBAs. Recent research reveals that the poor psychometric performance of WBAs may not be a result of divergent interpretations of the observed data, but rather of the questions and scales that were utilized. Many of the WBAs in EM now in use rating scales that are linked to a specified level of training (for example, "below, meets, or over expectations"). These scales, however, are based on the rater's anticipated performance for a specific training level. The upper end of the rating scale anchors used in entrustability measures are the standard of competence or autonomous performance. The way the descriptive anchors are referred to in the literature and how they are specifically phrased varies (e.g., entrustability, entrustment, and independence anchors) <sup>[16]</sup>.

### **Barriers to Establish WPBA in the Operating Theatre**

In order to achieve systematic supervised training in the operating room, the stakeholders have identified three levels of barriers:

- 1. Challenges at the Organizational Level-** By effectively advocating for better training conditions, such as allocating more theatre time per patient, ring-fencing beds for elective admissions, and establishing training opportunities at nearby diagnosis and treatment facilities, these may be amenable to change.
- 2. Professional Level Challenges:** These can be changed through proactive planning and workload reorganisation by the main stakeholders (clinical supervisors and trainees).
- 3. Individual Level Challenges -** These are amenable to direct change by particular clinical supervisory teams and trainee teams with the aim of improving their training environment, such as better matching of suitable trainees to suitable surgical cases <sup>[5]</sup>.

### **Methods to Improve WPBA as an Assessment Tool.**

Lack of suitable WBA tools does not pose a problem to implementation. WBA tools abound in our journals, each one created with an own set of elements, formats, procedures, and goals. Others require thorough assessment reports to be completed after months of supervision, while some only have a single item that must be filled after each activity. Some WBA have minimal or no stakes, while others have strict timelines for progression. Some people place a higher priority on the documentation of feedback than others, while many people try to balance the two. Therefore, rather than establishing "de novo" tools, the Accreditation Council for Graduate Medical Education (ACGME) advises referring at current WBA tools and evaluating their effectiveness when developing assessment procedures <sup>[61]</sup>.

Regular trainer education and training should be provided. This might also increase system trainer faith. A more enthusiastic, motivated, and engaged workforce is likely to result in better utilisation of WPBA. For optimal learning, trainees should be proactive in WBAs and inform their instructors of their plans to finish them well in advance. They should be committed, motivated trainees who engage in reflective practise. Lack of time is a major challenge that can be overcome by improved task organisation, scheduling WBA sessions in the schedule, and other methods.

To improve the validity and learning episode of "on the job learning," efforts should be made to include overall learning events, even if the outcome or experience was not favourable. The trainees should not be penalised based on the reflections. It is not recommended to utilise the individual WBAs as a summative measure <sup>[57]</sup>.

It became clear through stakeholder interactions with WBA's many answer formats that there is no one response format that works for all. Although a strong design may be appropriate for some teachers because it matches their method of feedback, other teachers may find it distracting and out of place. Additionally, the design could be excellent for specific circumstances at a particular time but need modifications for regional variations or developing changes in practise. The practicality is affected by teachers' preferences for various tools or their desire to utilise the tool that best fits a certain context. For example, an assessment programme that allows teachers to choose their tool may increase acceptability for teachers, which may compound the challenges of administering assessments to teachers and then gathering them for competency committees to make summative judgements for learners. A strategy worth looking into is the strategic use of variously developed instruments that best match the learning process of residents.

The coaching relationship between the trainer and trainee should be given priority, and attempts should be made to create WBA tools that may act as a mediator to enhance teaching, learning, and feedback sharing within that relationship at work. WBA design and procedure can mitigate coaching, but they also have the potential to damage the connection. It would be wise to keep an eye out for WBA procedures that would make it difficult for residents to request coaching and feedback if they felt it would add to the stress of completing forms <sup>[61]</sup>. The ability to directly examine what a trainee accomplishes in the workplace makes WBAs more appealing than other evaluations like multiple-choice questions or OSCEs. This reduces the need to infer competence from lower tiers of Miller's pyramid, hence elevating the validity of WBAs over that of other evaluations. The highest tier of Miller's pyramid is directly targeted. All parties involved should have enough training and assessment knowledge. To better comprehend the interactions between users, tools, and context—which in turn must be understood to further improve WBAs—it is essential to have a solid grasp of user views on WBAs. The transition from formative to summative WBAs causes trainees to put less emphasis on learning and more on performance, which lowers engagement with the feedback and compromises the validity of the WBAs. Additionally, because summative WBAs carry more weight, assessors are less strict with their comments and judgments, especially when they lack confidence in their ability to defend their choices. This frustrates assessors who are attempting to separate their responsibilities as teachers and assessors. Some have proposed that in order to address these problems, stakeholders should be informed of the intended purpose (whether formative or summative) of each WBA in order to promote responsible use. WBAs have generally been shown to have low reliability, while using entrustment-based scales to redefine scales increases dependability. For acceptable generalisability, around ten WBAs are needed. However, determining the "ideal number" of WBAs is challenging. Some claim that the number of WBAs needed should depend on how well trainees perform since for trainees who perform well, fewer WBAs are needed to produce accurate estimations of competence <sup>[39]</sup>.

In general, how WBAs are used (such as hasty retrospective completion) is influenced by users' attitudes toward, availability of time for, and training in WBAs, which reduces a

WBA's validity. Similar to this, leniency in the assessor-trainee connection can undermine validity or improve it by allowing longitudinal assessment. Results may vary depending on the environment, including the case's difficulty and speciality.

The amount of time needed for WBAs determines their viability. While their validity is greatly impacted by their intent, WBAs' validity can be improved by utilising entrustment-based definitions (i.e., summative or formative). The formative-summative tension, user disengagement, and assessor-trainee relationships are three significant topics that have an impact on the usability of WPBAs. These seem to be key factors to consider when building, implementing, and assessing WBAs <sup>[39]</sup>.

### **Role of Feedback in WPBA**

Feedback provides information and is not intended to be judgmental. The trainee is given specific, subjective comments on their observed performance in a way that is useful for them to consider and use to improve their future performance. Providing good quality and timely feedback has an essential role in learning and professional development in medicine. Feedback during learning allows students to take feedback on board immediately and to try to realise improvement during the learning process. This is often more effective and productive to the learning experience than end-of task feedback measures (usually summative) which require students to remember the feedback and apply the recommended strategies to a future task. There are several methods described to help assessors to provide feedback to trainees. One of the older but more commonly used feedback techniques in clinical medicine is that described by Pendleton <sup>[62]</sup>. Pendleton's rules of feedback include the following:

1. The trainee identifies what went well and discusses what did not go well and how they could improve this aspect of performance.
2. The assessor highlights what they observed went well and identifies observed areas for improvement.
3. Both agrees the areas of improvement and works on the action plan for future improvement.

The following rules to be followed <sup>[63]</sup> for the feedback to be effective.

1. Be clear
2. Be specific
3. Be constructive
4. Be descriptive rather than evaluative
5. Timing is important
6. Comment on the behaviour and not the personality

Teachers perceived that their primary responsibility is mentoring students to develop their surgical skills (i.e., for formative feedback). The teachers all expressed a desire to aid residents in honing their surgical abilities and preparing for solo practise. They stated a need for assessment tools to promote formative feedback because they stated that feedback was their purpose for assessment. Numerous educators underlined the need for a pathway for remediation whenever negative, critical, or constructive feedback is given. This will allow students to advance through more coaching and clinical experiences <sup>[61]</sup>.

Narrative feedback is the most helpful because it is more relevant and easier to understand than numerical ratings. Feedback should be observation-based and focus on certain aspects of the interaction. Feedback should also be concentrated on general procedures to promote transferability, be actionable, and ideally identify goals tailored to trainees' learning objectives and needs in a timely manner to maximise its relevance and trainees' opportunities to put it into practise. It should also be benchmarked to be developmentally appropriate, honest, and framed using standards or norms to provide trainees with clear benchmarks to work toward. In order to give senior trainees a comprehensive picture of their progress, holistic feedback is especially valuable. This can be helped by compiling input from many assessors. Consistencies and a preference for focusing on individual characteristics, such as mannerisms, were themes of low-quality comments <sup>[39]</sup>.

MSF is thought to boost employee motivation, resulting in positive behaviour changes, higher output, and increased self-awareness, all of which are vital for the progress of any organisation <sup>[64]</sup>. The Sheffield Peer Review Assessment Tool (SPRAT), used by MSF, does not give enough information on trainees about whom concerns are raised, according to Archer et al.'s non-comparative action-based study <sup>[65]</sup>. More assessments are needed for these trainees. Additionally, they believed that unchecked assessor self-selection introduces leniency bias and that it should be stopped. According to qualitative research by Sargeant et al, only few doctors who received negative feedback from MSF modified their behaviour <sup>[66]</sup>. Doctors did not alter their behaviour if the feedback was positive. The most common critique was particular, related to communication skills, and was provided by patients. The area of feedback used the least by medical colleagues was clinical competence. In a different qualitative study using focus groups and interviews, Sargeant et al. found that family doctors typically agreed with the comments from their patients <sup>[66]</sup>. In a prospective observational cohort study, Holmboe et al. collected feedback from mini-CEX sessions and demonstrated that mini-CEX often yields an improvement suggestion, with the majority of the suggestions focusing on the clinical abilities of medical interviewing, physical examination, and counselling <sup>[67]</sup>. Marriot et al study on PBA revealed that trainees rated the clinical supervisor's input as ranging from somewhat to very helpful. Feedback was assessed similarly by clinical supervisors <sup>[68]</sup>. According to Canavan et al., many MSF forms had no comments at all, and of those that did, a significant portion were found to lack information that may be utilised to act, decreasing the utility of such comments <sup>[69]</sup>. In the study by Burford et al., most of the trainees did not anticipate changing their behaviour in response to feedback from the MSF tools utilised, although the TAB's perceived utility was consistently greater than the mini-PAT's <sup>[70]</sup>.

Left unchecked, feedback may have significant and, in some circumstances, devastating repercussions for trainees who are "at danger of failing," "underperforming," or "in difficulty." In the lack of precise performance metrics, attempts to define "underperformance" or "bad performance" remain largely subjective. The most restrictive (2013) definition given in research conducted in the UK is that a trainee who is underperforming "requires intervention above the regular degree of supervisor-trainee engagement". This is a descriptive definition, but it does not identify the main reason why the trainee is having problems; rather, it gives a broad description of a trainee who is not yet reaching the requirements of their training level. Aileen Barrett from Ireland conducted a systematic review Using pre-established, internationally recognized, BEME (Best Evidence in Medical Education) Collaboration guidelines to address the following research questions:

1. Can workplace-based assessment be used to identify and remediate underperformance among postgraduate medical trainees?
2. Of those tools thought to identify and/or remediate underperforming trainees, what features specifically contribute to their usefulness for identifying or remediating underperformance among postgraduate medical trainees?

The authors found that Trainees who had been judged to be underwhelming or performing poorly (by other measures) did not always select less complex cases for their WBA. To complete a direct observation of procedural skills (DOPS) evaluation and conduct a mini-PAT, this group of trainees was more likely to approach a nursing colleague. This would suggest that those who were aware of their performance issues avoided their senior colleagues and medical peers to some extent <sup>[71]</sup>. A change in practise may be less likely to occur and assessments may become more of a "tick-box exercise" if a competency or element of performance is found to be "achieving" or "above expectation," according to the potential "ceiling effect" of WBA rating systems <sup>[72]</sup>. It is crucial to fully investigate the tools' capabilities to find the weaker performance baseline and/or to help raise performance above this baseline.

### III. CONCLUSION

The workplace-based assessments bring higher levels of learning as it touches higher levels of learning on Bloom's taxonomy, i.e., applying, analysing, and evaluating. It is a low stake assessment used for formative assessment. Since they are designed to be performed more frequently and are primarily intended for low-stakes formative purposes (assessment for learning), they may not be as reliable as assessments that are primarily intended for summative purposes (assessment of learning), which are intended to make high-stakes decisions about certification or training progression. But it can be used as a summative tool if the reliability of the WPBA assessment is increased. Since trainees are evaluated based on direct observation of their actual clinical practise, WPBAs have high face validity as measurements of performance on a day-to-day basis. Even though the WPBAs are being used then and there in postgraduate students to assess their competency, it is not included as a part of postgraduate undergraduate curriculum except a few courses like foundation course in UK postgraduate education. But it should be incorporated in regular UG curriculum as a formative assessment especially in CBME curriculum to assess the competency in the workplace as an authentic assessment.

WPBAs can be implemented by giving clinical supervisors and trainees useful, timely training on the tools, encouraging appropriate tool use for formative assessment, demonstrating in practise the viability of workplace learning and assessment, and conducting "field testing" that prompts tool modification. By providing assessors and trainees with continual training and support, conflicts and prejudices may be reduced. Additionally, if many assessors are employed, caution must be taken to prevent rating contamination by deferring required discussion until after ratings have been issued.

## REFERENCES

- [1] Yuying Liang and Lorraine M. Noble. Chinese doctors' views on workplace-based assessment: trainee and supervisor perspectives of the mini-CEX. *Medical education online* 2020, VOL. 26, 1869393. <https://doi.org/10.1080/10872981.2020.1869393>
- [2] Govaerts M. Workplace-based assessment and assessment for learning: threats to validity. *J Grad Med Educ.* 2015;7(2):265–7. <https://doi.org/10.4300/JGME-D-15-00101.1>.
- [3] Miller GE. The assessment of clinical skills/ competence/ performance. *Acad Med.* 1990;65:S63–67.
- [4] Pulito, Andrew & Donnelly, Michael & Plymale, Margaret & Mentzer, Robert. (2006). What Do Faculty Observe of Medical Students' Clinical Performance?. *Teaching and learning in medicine.* 18. 99-104. [10.1207/s15328015tlm1802\\_2](https://doi.org/10.1207/s15328015tlm1802_2).
- [5] JD Beard, J Marriott, H Purdie and J Crossley. Assessing the surgical skills of trainees in the operating theatre: a prospective observational study of the methodology. *Health Technology Assessment* 2011; Vol. 15: No. 1
- [6] Andrea C. Lörwald, Felicitas-Maria Lahner, Robert Greif, Christoph Berendonk, John Norcini & Sören Huwendiek (2017): Factors influencing the educational impact of Mini-CEX and DOPS: A qualitative synthesis, *Medical Teacher*, DOI: 10.1080/0142159X.2017.1408901
- [7] Ferguson PC, Caverzagie KJ, Nousiainen MT, Snell L. Changing the culture of medical training: an important step toward the implementation of competency-based medical education. *Med Teach.* 2017;39(6):599–602
- [8] Frank JR, Snell LS, Cate OT, Holmboe ES, Carraccio C, Swing SR, et al. Competency-based medical education: theory to practice. *Med Teach.* 2010;32(8):638–45.
- [9] Yousuf Guraya S. Workplace-based Assessment; Applications and Educational Impact. *Malays J Med Sci.* 2015 Nov;22(6):5-10. PMID: 28223879; PMCID: PMC5295751
- [10] Weston PS, Smith CA. 2014. The use of mini-CEX in UK foundation training six years following its introduction: lessons still to be learned and the benefit of formal teaching regarding its utility. *Med Teach.* 36:155–163
- [11] Wall D, Singh D, Whitehouse A, Hassell A, Howes J. Self-assessment by trainees using self-TAB as part of the team assessment of behaviour multisource feedback tool. *Med Teach.* 2012;34(2):165–167. doi: 10.3109/0142159X.2012.644840.
- [12] Campbell J, Richards S, Dickens A, Greco M, Narayanan A, Brearley S. Assessing the professional performance of UK doctors: an evaluation of the utility of the General Medical Council patient and colleague questionnaires. *Qual Saf Health Care.* 2008;17(3):187–193. doi: 10.1136/qshc.2007.024679.
- [13] Mackillop LH, Crossley J, Vivekananda-Schmidt P, Wade W, Armitage M. A single generic multi-source feedback tool for revalidation of all UK career-grade doctors: Does one size fit all? *Med Teach.* 2011;33(2):e75–e83. doi: 10.3109/0142159X.2010.535870.
- [14] Marcio M. Gomes et al. Teaching and assessing intra-operative consultations in competency-based medical education: development of a workplace-based assessment instrument. *Virchows Archiv* (2021) 479:803–813. <https://doi.org/10.1007/s00428-021-03113-6>
- [15] Michael Gottlieb et al. Direct Observation Tools in Emergency Medicine: A Systematic Review of the Literature. doi: 10.1002/aet2
- [16] Cheung et al. The Ottawa Emergency Department Shift Observation Tool (O-EDShOT): A New Tool for Assessing Resident Competence in the Emergency Department. doi: 10.1002/aet2.10419
- [17] Bandiera G, Lendrum D. Daily encounter cards facilitate competency-based feedback while leniency bias persists. *Can J Emerg Med* 2008;10:44–50
- [18] Sherbino J, Kulasegaram K, Worster A, Norman GR. The reliability of encounter cards to assess the CanMEDS roles. *Adv Heal Sci Educ* 2013;18:987–96.
- [19] Harden RM, Gleeson FA. Assessment of clinical competence using an objective structured clinical examination (OSCE) *Med Educ.* 1979;13:38–54
- [20] Norcini JJ, Blank LL, Arnold GK, Kimball HR. The mini-CEX (clinical evaluation exercise): a preliminary investigation. *Ann Intern Med.* 1995;123(10):795–9. doi:10.1059/0003-4819-123-10-199511150-00008.
- [21] Wilkinson JR, Crossley JG, Wragg A, Mills P, Cowan G, Wade W. Implementing workplace-based assessment across the medical specialties in the United Kingdom. *Med Educ.* 2008 Apr;42(4):364-73. doi: 10.1111/j.1365-2923.2008.03010.x. PMID: 18338989.
- [22] Carr SJ. Assessing clinical competency in medical senior house officers: how and why should we do it? *Postgrad Med J.* 2004 Feb;80(940):63-6. doi: 10.1136/pmj.2003.011718. PMID: 14970290; PMCID: PMC1742939.

- [23] Rethans JJ, Norcini JJ, Barón-Maldonado M, Blackmore D, Jolly BC, LaDuca T, Lew S, Page GG, Southgate LH. The relationship between competence and performance: implications for assessing practice performance. *Med Educ.* 2002 Oct;36(10):901-9. doi: 10.1046/j.1365-2923.2002.01316.x. PMID: 12390456.
- [24] Sloan DA, Donnelley MB, Schwartz R, et al. The objective structured clinical examination—the new gold standard for evaluating postgraduate clinical performance. *Ann Surg* 1995;222:735–42.
- [25] Novack DH, Cohen D, Peitzman SJ, et al. A pilot test of webOSCE: a system for assessing trainees' clinical skills via teleconference. *Medical Teacher* 2002;24:483–7
- [26] Issenberg SB, McGaghie WC, Waugh RA. Computers and evaluation of clinical competence. *Ann Intern Med* 1999;130:244
- [27] Klass D. Revaluation of clinical competency. *Am J Phys Med Rehabil* 2000;79:481–6
- [28] Boulet JR, McKinley DW, Norcini JJ, et al. Assessing the comparability of standardised patient and physician evaluations of clinical skills. *Advances in Health Sciences Education* 2002;7:85–97
- [29] Barrett A, Galvin R, Scherpbier AJJA, et al. *Postgrad Med J* 2017;93:138–142
- [30] Wu Y, Gong M, Zhang D, Zhang C. Educational impact of the mini-Clinical Evaluation Exercise in resident standardization training: a comparative study between resident and professional degree postgraduate trainees. *J Int Med Res.* 2020 May;48(5):300060520920052. doi: 10.1177/0300060520920052. PMID: 32459121; PMCID: PMC7278105
- [31] Xu D, Sun B, Wan X, et al. Reformation of medical education in China. *Lancet* 2010; 375: 1502–1504
- [32] DeLisa J A. Evaluation of clinical competency. *Am J Phys Med Rehabil* 2000;79:474–477. [PubMed] [Google Scholar]
- [33] Durning S J, Cation L J, Markert R J. *et al* Assessing the reliability and validity of the mini-clinical evaluation exercise for Internal medicine residency training. *Acad Med* 2002;77:900–904.
- [34] Mitchell C, Bhat S, Herbert A, Baker P. Workplace-based assessments of junior doctors: do scores predict training difficulties? *Med Educ.* 2011 Dec;45(12):1190-8. doi: 10.1111/j.1365-2923.2011.04056.x. Epub 2011 Oct 13. PMID: 21995509
- [35] McKavanagh P, Smyth A, Carragher A. Hospital consultants and workplace based assessments: how foundation doctors view these educational interactions? *Postgrad Med J.* 2012 Mar;88(1037):119-24. doi: 10.1136/postgradmedj-2011-130121. Epub 2012 Jan 13. PMID: 22247316
- [36] McLeod RA, Mires GJ, Ker JS. The use of the Direct observation of procedural skills (DOPS) assessment tool in the clinical setting - The perceptions of students. *Int J Clin Ski.* 2011;5:102–7
- [37] Morris A, Hewitt J, Roberts CM. Practical experience of using directly observed procedures, mini-clinical evaluation examinations, and peer observation in pre-registration house officer (FY1) trainees. *Postgrad Med J.* 2006;82:285–8.
- [38] Jason Ali and Aaron Goh. Student perceptions of workplace-based assessment. *The clinical teacher* 2016; 13: 1–6
- [39] Shaun Prentice, Jill Benson, Emily Kirkpatrick, Lambert Schuwirth. Workplace-based assessments in postgraduate medical education – a hermeneutic review. doi: 10.1111/MEDU.14221
- [40] Andreou V, Peters S, Eggermont J, Embo M, Michels NR, Schoenmakers B. Fitness-for-purpose of the CanMEDS competencies for workplace-based assessment in General Practitioner's Training: a Delphi study. *BMC Med Educ.* 2023 Apr 1;23(1):204. doi: 10.1186/s12909-023-04207-2. PMID: 37005633; PMCID: PMC10067520.
- [41] Frank JR, Snell L, Sherbino J. CanMEDS 2015 physician competency framework: Royal College of Physicians & Surgeons of Canada. 2015.
- [42] Norman G. CanMEDS and other outcomes. *Adv Health Sci Educ.* 2011;16(5):547–51.
- [43] Nousiainen MT, Caverzagie KJ, Ferguson PC, Frank JR. Implementing competency-based medical education: What changes in curricular structure and processes are needed? *Med Teach.* 2017;39(6):594–8
- [44] Nesbitt A, Baird F, Canning B, Griffin A, Sturrock A. Student perception of workplace-based assessment. *Clin Teach* 2013;10(6):399–404
- [45] Al-Kadri HM, Al-Kadi MT, Van Der Vleuten CP. Workplace-based assessment and students' approaches to learning: a qualitative inquiry. *Med Teach* 2013;35:S31–S38
- [46] Kamat C, Todakar M, Patil M, Teli A. Changing trends in assessment: Effectiveness of Direct observation of procedural skills (DOPS) as an assessment tool in anesthesiology postgraduate students. *J Anaesthesiol Clin Pharmacol.* 2022 Apr-Jun;38(2):275-280. doi: 10.4103/joacp.JOACP\_329\_20. Epub 2022 Jul 28. PMID: 36171945; PMCID: PMC9511833.
- [47] Hill K, Silcock D, Burns F, Jeffrey S, Chaudhri S. Abstract PR071: Use of a directly observed procedural skills assessment tool for candidate evaluation after an ultrasound guided central line insertion workshop. *Anesth Analg.* 2016;123:101–2

- [48] Kumar N, Singh NK, Rudra S, Pathak S. Effect of formative evaluation using direct observation of procedural skills in assessment of postgraduate students of obstetrics and gynaecology: Prospective study. *J Adv Med Educ Prof.* 2017;5:1–5.
- [49] Barton JR, Corbett S, van der Vleuten CP. English Bowel Cancer Screening Programme; UK Joint Advisory Group for Gastrointestinal Endoscopy. [i]The validity and reliability of a direct observation of procedural skills assessment tool: Assessing colonoscopic skills of senior endoscopists. *Gastrointest Endosc.* 2012;75:591–7
- [50] Ali L, Ali S, Orakzai N, Ali N. Effectiveness of Direct observation of procedural skills (DOPS) in postgraduate training in urology at institute of kidney diseases, Peshawar. *J Coll Physicians Surg Pak.* 2019;29:516–9.
- [51] Profanter C, Perathoner A. DOPS (Direct Observation of Procedural Skills) in undergraduate skills-lab: Does it work? Analysis of skills-performance and curricular side effects. *GMS Z Med Ausbild.* 2015;32:Doc45
- [52] Joshi MK, Singh T, Badyal DK. Acceptability and feasibility of mini-clinical evaluation exercise as a formative assessment tool for workplace-based assessment for surgical postgraduate students. *J Postgrad Med.* 2017 Apr-Jun;63(2):100-105. doi: 10.4103/0022-3859.201411. PMID: 28272063; PMCID: PMC5414419.
- [53] Khalil S, Aggarwal A, Mishra D. Implementation of a Mini-Clinical Evaluation Exercise (Mini-CEX) Program to Assess the Clinical Competence of Postgraduate Trainees in Pediatrics. *Indian Pediatr.* 2017 Apr 15;54(4):284-287. doi: 10.1007/s13312-017-1089-z. Epub 2017 Feb 2. PMID: 28159950
- [54] Gupta S, Sharma M, Singh T. The Acceptability and Feasibility of Mini-clinical Evaluation Exercise as a Learning Tool for Pediatric Postgraduate Students. *Int J Appl Basic Med Res.* 2017 Dec;7(Suppl 1):S19-S22. doi: 10.4103/ijabmr.IJABMR\_152\_17. PMID: 29344452; PMCID: PMC5769164.
- [55] Sethi S, Srivastava V, Verma P. Mini-Clinical Evaluation Exercise as a Tool for Formative Assessment of Postgraduates in Psychiatry. *Int J Appl Basic Med Res.* 2021 Jan-Mar;11(1):27-31. doi: 10.4103/ijabmr.IJABMR\_305\_20. Epub 2021 Jan 26. PMID: 33842292; PMCID: PMC8025951
- [56] Batra P, Batra R, Verma N, Bokariya P, Garg S, Yadav S. Mini clinical evaluation exercise (Mini-CEX): A tool for assessment of residents in department of surgery. *J Educ Health Promot.* 2022 Aug 25;11:253. doi: 10.4103/jehp.jehp\_1600\_21. PMID: 36325223; PMCID: PMC9621368.
- [57] Kamal Raj Aryal, et al. Work-Based Assessments in Higher General Surgical Training Program: A Mixed Methods Study Exploring Trainers' and Trainees' Views and Experiences. *Surg J* 2020;6:e49–e61
- [58] Damian J et al. How Trainees Come to Trust Supervisors in Workplace-Based Assessment: A Grounded Theory Study. *Acad Med.* 2022;97:704–710.
- [59] Bourdieu P. *The Logic of Practice.* Cambridge, UK: Polity Press; 1990
- [60] Patel P, Martimianakis MA, Zilbert NR, et al.. Fake it 'til you make it: Pressures to measure up in surgical training. *Acad Med.* 2018;93:769–774
- [61] Nawaaz A. Nathoo, Ravi Sidhu & Andrea Gingerich (2020): Educational Impact Drives Feasibility of Implementing Daily Assessment in the Workplace, Teaching and Learning in Medicine, DOI: 10.1080/10401334.2020.1729162
- [62] Pendleton D, Schofield T, Tate P. A method for giving feedback. In: *The consultation: an approach to learning and teaching.* Oxford: Oxford University Press, 1984:68–71
- [63] Branch W T, Paranjape A. Feedback and reflection: teaching methods for clinical settings. *Acad Med* 2002;77:1185–1188
- [64] Abdulla A: A critical analysis of mini peer assessment tool (mini-PAT). *J R Soc Med* 2008, 101:22–26.
- [65] Brown, Jeremy & Lowe, Kathryn & Fillingham, Jill & Murphy, Philip & Bamforth, Margaret & Shaw, N j. (2014). An investigation into the use of multi-source feedback (MSF) as a work-based assessment tool. Medical teacher. 36. 10.3109/0142159X.2014.909920.
- [66] Sargeant J, Mann K, Sinclair D, van der Vleuten C, Metsemakers J. Challenges in multisource feedback: intended and unintended outcomes. *Med Educ.* 2007 Jun;41(6):583-91. doi: 10.1111/j.1365-2923.2007.02769.x. PMID: 17518839.
- [67] Holmboe ES, Yepes M, Williams F, Huot SJ. Feedback and the mini clinical evaluation exercise. *J Gen Intern Med.* 2004 May;19(5 Pt 2):558-61. doi: 10.1111/j.1525-1497.2004.30134.x. PMID: 15109324; PMCID: PMC1492325.
- [68] Marriott J, Purdie H, Crossley J, Beard JD: Evaluation of procedure-based assessment for assessing trainees' skills in the operating theatre. *Br J Surg* 2010, 98(3):450–457
- [69] Canavan C, Holtman MC, Richmond M, Katsufakis PJ: The quality of written comments on professional behaviors in a developmental multisource feedback program. *Acad Med* 2010, 85(10 Suppl):S106–S109.



- [70] Burford B, Illing J, Kergon C, Morrow G, Livingston M: User perceptions of multi-source feedback tools for junior doctors. *Med Educ.* 2010, 44 (2):165–76. Epub 2010 Jan 5.
- [71] Barrett et al. A BEME (Best Evidence in Medical Education) systematic review of the use of workplace-based assessment in identifying and remediating poor performance among postgraduate medical trainees. *Systematic Reviews* (2015) 4:65
- [72] Archer J, McGraw M, Davies H: Republished paper: Assuring validity of multisource feedback in a national programme. *Postgrad Med J* 2010, 86 (1019):526–31
- [73] Goff, B.A., Nielsen, P.E., Lentz, G.M. et al. (2002). Surgical skills assessment: a blinded examination of obstetrics and gynecology residents. *American Journal of Obstetrics and Gynecology* 186 (4): 613–617
- [74] Maatsch JL, Krome RL, Sprafka S, Maclean CB. The Emergency Medicine Specialty Certification Examination (EMSCE) *JACEP.* 1976;5(7):529–534.
- [75] <https://www.rcpath.org/static/bf959429-aa0c-48d7-a267902f7c8609dd/definitions-of-assessment-tools.pdf>
- [76] Burkill G. Work-based assessment for trainees—more than just a few new tools? *Clin Radiol.* 2008;63(1):12–14. doi: 10.1016/j.crad.2007.07.014
- [77] Cheung WJ, Dudek N, Wood TJ, Frank JR. Daily Encounter Cards-Evaluating the Quality of Documented Assessments. *J Grad Med Educ.* 2016 Oct;8(4):601-604. doi: 10.4300/JGME-D-15-00505.1. PMID: 27777675; PMCID: PMC5058597.