

Invited Article

MEDICAL EDUCATION

Challenges and strategies in implementing workplace-based assessments in the Sri Lankan medical education system

Gitanjali Sathiadas

Faculty of Medicine, University of Jaffna, Sri Lanka

Keywords: Work Place Based Assessments, WPBA, competency

Corresponding Author: Gitanjali Sathiadas, E-mail: < mgsathiadas@gmail.com >  <https://orcid.org/0000-0002-7822-0447>
Received: May 2018, Accepted revised version: JSeptember 2018, Published: September 2018
Competing Interests: Author has declared that no competing interests exist

© Author. This is an open-access article distributed under a Creative Commons Attribution-Share Alike 4.0 International License (CC BY-SA 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are attributed and materials are shared under the same license.



Introduction

Assessment is an intrinsic part of the training of a medical professional. In the early period, training in medicine followed the apprenticeship model, where the learner observed, assisted and performed in the clinical setting and the mentor gave feedback and training. Subsequently, the training became more compartmentalised with preclinical and clinical stages. Assessment methods also conformed to this curriculum plan. Evolution of the medical curriculum has shifted the focus to learning while doing, the highest level of the Miller's pyramid. Assessments, in turn, are moving away from high stakes examinations towards gathering evidence of clinical competence and professional behaviour in the workplace. The complexity of medical competencies is such that a single method may not be able to holistically assess one competency leading to the use of multiple assessment tools. A clear standard is needed, and performance below that standard is considered as not fit to practice. Norm referencing is not acceptable in the assessment of clinical competencies

Assessments may be formative or summative. When a benchmark is set during formative assessments, it can reinforce students' intrinsic motivation to learn and inspire them to set higher standards for themselves. Summative assessments, on the other hand, are focused mainly on professional self-regulation and ensuring accountability; they may also act as a barrier to further practice or training. [1] Even though summative assessment do not involve feedback to drive learning, they may still influence learning as students tend to learn what they expect to be tested on [1]. There are several methods of testing in both formative and summative assessments.

Assessments are judged by their reliability (the degree to which the measurement is accurate and reproducible), validity (whether the assessment measures what it claims to measure), impact on future learning and practice, acceptability to learners and faculty

and cost (to the individual trainee, the institution, and society at large) [2]. Assessments become more valid and reliable when the type of assessment is varied, the assessors are different, and the assessment is conducted over a period of time. Hence a single method cannot cover all aspects of competencies of the layers of Miller's pyramid and a combination of methods is needed. Table 1 shows the reliability of different tests versus test duration. In general, the longer the time taken to do the assessment the more reliable the result

Table 1: Estimated reliability scores of different test methods

Instrument	Reliability for different testing times			
	1 hour	2 hours	4 hours	8 hours
MCQs*	0.62	0.76	0.93	0.93
Patient management problem*	0.36	0.53	0.69	0.82
Key feature case – case vignette*	0.32	0.49	0.66	0.79
Oral examination ⁺	0.50	0.69	0.82	0.90
Long case ⁺	0.60	0.75	0.86	0.90
Mini CEX [@]	0.73	0.84	0.92	0.96
OSCE*	0.54	0.69	0.82	0.90
Video based patient encounter ⁺	0.62	0.76	0.93	0.93
Standard patient encounter [@]	0.61	0.76	0.82	0.86

MEDICAL EDUCATION 2005; 39: 309–317

*One-facet all random design with items crossed with patients (pxi)

⁺Two-facet all random design with judges(examiners) nested within items within persons

[@]One-facet all random design with items nested within persons

Several studies have shown discrepancies between what doctors can do in high stake-controlled examination situations to the actual practice situation. [3] It is understood that development of many competencies will lead to a certain level of performance in an actual situation. This is the rationale for introducing workplace-based assessments (WPBA) [4]. WPBAs are used as a tool to measure movement up the Miller's pyramid from performance based (in "shows how") to competence based (in "does how") as well as ensure assessment of practice performances [3]. The assessment tools have to be reliable, valid and feasible to be implemented in the practice setting [5].

Educational basis of WPBA

Experts believe that assessments of actual practice give a much better reflection of competence when compared to assessment under test conditions. In Miller's framework of assessments performance (shows how) and action (does) are at the highest level of the pyramid as they collate doctors' performance in everyday practice rather than in an artificial exam situation. WPBA should be part of a structured training programme that should include induction, systematic teaching, WPBA and ongoing feedback and encourage a holistic approach with reflective practice and lifelong learning. Junior doctors/trainees should be asked to complete a certain number of assessments on the

job and these assessments should be reviewed at the end of each training year from a portfolio of ongoing WPBAs. [4,5]

WPBA lead to a deeper approach to learning independent of any summative effect [6]. WPBAs foster self-directed learning which is essential for continuing professional development. Clinical encounters relevant to the trainee can be assessed at the workplace and interaction between the assessor and trainee provides a life time learning experience [5].

It is stated that WPBAs cannot replace traditional methods of assessments but carry a potential for add-on methods, especially to the in-training and formative assessments. [7]. WPBAs can compensate for some shortcomings in traditional assessment methods. They align with learning and actual working and provide a good ground for reflection which is very important in the learning process [4].

Clinical skills are not restricted to history taking and examination but include several soft skills such communication skills, professionalism, and ethics, also referred to as the non-cognitive component of clinical skills. Hence it is important to include a formal training and an effective assessment of these soft skills, along with the technical clinical skills, in the medical curriculum [4].

One of the main educational aims of WPBA is competency. Competence is not an achievement, but a process of lifelong learning and it plays an integral role in helping physicians identify and respond to their own learning needs [4]. An assessment of competence must provide an insight into actual performance, capacity to adapt to change, capacity to generate new knowledge from the existing knowledge and capacity to improve overall performance.

Evidence based research has shown that systematic feedback delivered by the supervisor enhances clinical performance [8]. Effective feedback on the performance of the trainee will thereby enhance the learning process.

Methods available

There are several assessment tools aiming to assess various facets of the trainee's performance. Table 2 summarises the categories.

Table 2: Categories of work place-based assessments

No	Task	Tool
1	Direct observation – single encounter	Mini-Clinical Evaluation Direct Observation of Practical Procedural skills Case based discussion
2	Multiple encounters	Multi source feedback Patient satisfaction questionnaire
3	Aggregation and reflection measures	Log book Portfolio

Mini-Clinical Evaluation Exercises (MiniCEX)

This is a 15-minute snapshot of a doctor-patient encounter, to assess clinical skills, attitudes and behaviour essential for high quality patient care. It permits evaluation based on the clinical setting and patient problems and is administered on site. Trainees are encouraged to choose a different assessor for each assessment covering areas such as history taking, physical examination, professionalism, clinical judgement, communication skills, organisation, efficiency and overall clinical care. The trainee is rated on a nine-point Likert scale and given a final rating as well. For each speciality, the number of MiniCEXs may vary and a minimum of four to six per year is encouraged [9].

The strengths of MiniCEXs are that they can be used in different clinical settings like clinics, ward rounds and on calls. Each trainee being evaluated by different assessors in much broader clinical situations gives the trainee an opportunity to get feedback from more than one faculty member. Training of the assessor on effective feedback can ensure reproducibility of the assessments. These assessments cause less anxiety to the trainee when compared to traditional formats as they are less formal and less dependent on one encounter. The weakness of MiniCEXs are that they have to be scheduled and are time consuming and trainees are also not observed for the duration of the complete patient encounter.

The validity of assessment is vital in such circumstances. When used in isolation, psychometric analyses of the construct may not provide the evidence of validity and educational impact. Hence literature suggests the use of Longitudinal Evaluation of Performance (LEP) [6]. This method gives the trainer a fixed reference point to grade the trainee as 'need improvement', 'satisfactory' and 'superior performance'. Being primarily a formative assessment, if the LEPs indicate a need for improvement, it will only be a requirement to show progress in subsequent encounters and show satisfactory completion.

The MiniCEXs are mainly screening assessments that can identify personal strengths and weaknesses. Educational impact of the process and enhancing feedback are maximised during these encounters [6].

Directly Observed Practical Procedures (DOPS)

This is specifically designed to assess procedural skills involving real patients in a single encounter. In this method of assessment, the trainer observes the trainee when he/she does a practical procedure, rates it and gives feedback on it. This is a valuable learning opportunity for trainees to enhance performance in a skill. This method of assessment has been shown to be valid, reliable and feasible in evaluating postgraduate doctors in the UK [10].

The number of encounters required varies with the speciality. For simpler procedures, mainly relevant to junior trainees (e.g. cannulation), this may not be an issue. On the other hand, for more senior trainees performing more complex procedures it may be difficult and frustrating to achieve adequate numbers of DOPS assessments [11].

Case Based Discussion (CBD)

This is a structured discussion between the trainee and the educational supervisor about how a case was managed by the trainee, talking through the reasons for the actions that have been recorded. The discussion must be around the actual case and not on hypothetical events. The assessor explores the trainee's clinical judgment and professional behaviour. Discussion can be based on trainee's diagnostic and management skills. Assessor then provides constructive feedback to the trainee. CBDs can explore a full range of holistic and complex situations, the course of action, explaining the course of action and reflecting on the outcome. This can test higher order thinking and how trainees prioritise applied knowledge. It also gives an opportunity to talk about record keeping and presentation skills.

CBDs have significant face and content validity and good reliability [9]. Validity and reliability mainly depend on the assessor's training [10]. The number of encounters needed per year is around 4-6 for it to be valid [9].

Multisource feedback (360° assessment)

It is understood that the medical curriculum is not limited to knowledge and skills but also includes behaviour and values. Multisource feedback (MSF) has emerged as an effective tool to assess professional attitude and behaviour in the workplace. It can heighten team-working, productivity, good communication and trust. This method is practical, valid and reliable [12]. MSF is a model of assessment in which a number of colleagues act as assessors of an individual. Their assessments are recorded on paper or electronically and the observations are fed back to the individual, either directly or with the help of a mentor or supervisor. Assessment measures can vary in terms of: number of assessors; method of assessor selection and the content of the pro forma used in the assessment [12]. As these are done as multi-rater assessments MSF has been shown to be a valid and useful tool [12].

The number of assessors needed for reliable results depend on the content and goal of the MSF, the number of items included in the questionnaire, the competencies assessed and the assessors' training. For instance, for the assessment of interpersonal skills and professionalism, more assessors are needed [13]. Ramsey and colleagues (1996) stated that 10 to 11 responses per physician were necessary to achieve a generalizability coefficient of 0.7. Wood & Campbell (2004) calculated the need for eight raters of Obstetrics and Gynaecology trainees to give a coefficient of 0.8 [14,15].

MSF can be used to develop an insight into the strengths and weaknesses of an individual, to enhance a culture change, for summative assessment, to evaluate the potential of individuals for career advice/selection, to enhance the effectiveness of a team and to identify the training needs of the programme [12]. In almost all published data on MSF it is stated that the main purpose is to identify those who may have a problem in the interpersonal domain [12]. Although some may feel embarrassed by negative feedback or feel aggressive towards the system, it rescues poorer trainees, protects patients, reduces opportunity for disharmony and complaints and provides evidence which may be useful in the future [12].

Challenges encountered in conducting WPBA

1. **Feasible sample to achieve reliable inferences:** Studies have been conducted to assess the number of direct observations required to get a reliable inference of the candidate. Despite variation between studies a reliable inference can be made with very feasible samples, between 8 and 10, irrespective of the type of instrument and of what is being measured [0].
2. **Bias:** Most WPBA are based on global judgements and are prone to bias. Bias in direct observation methods occurs as inflation of the scores. MSF bias is seen in selection of assessors, assessment context, relationship with the learner and the feedback process. Bias can occur during the self-assessment process, too, as we can either underestimate or overestimate our capabilities.
3. **Validity, reliability and feasibility of the instrument:** A valid instrument has to be designed and must be suitable for the situation that is assessed. It is also felt that validity is more dependent on the user than the instrument itself [14]. The instrument should be based more on qualitative narrative information rather than rating scales. Reliability is also dependent on the trainer and the training to use these tools. As WPBA is not a complicated process feasibility is relatively good [17].
4. **Effectiveness:** In a culture that adheres strictly to summative assessments, the value of WPBAs can be disregarded. Most WPBAs are formative assessments and are not combined with the summative assessment in the local setup. Combining the portfolio with the summative assessment will ensure triangulation as well as give a purpose to the portfolio [18].
5. **Time:** Lack of time has been a major barrier to the WPBA process, for both trainees and trainers. Scheduling the WPBA to the end of an attachment has been the common practice. WPBA should be included in the timetable and scheduled during the course.
6. **Training:** Training is essential to both the trainer and the trainee, mainly in reflection and the feedback process. The assessor's grade and qualifications also matter. Special attention should be paid on training in effective feedback. Positive feedback was perceived well by the students whereas it was difficult for students to accept direct criticism and negative feedback. Influence on students' performance had an undesirable effect due to negative feedback. Students also became hesitant to study and to interact with their supervisors in order to avoid another negative comment. This led to patchy reading, resulting in a tendency to surface learning. These facts have to be kept in mind during training, especially when giving negative feedback [19].
7. **Data protection and confidentiality:** Data should be protected when giving feedback and during assessment, in whatever method it is done, paper based or electronic based.
8. **Rules and regulations of the universities:** Most universities have regulations to record and assess students and trainees with summative assessments to grade them. The administration has to adapt to change with modern trends in education.

Potential assessment strategies to overcome the challenges of WPBAs

It is best to expose the student to WPBAs at medical school itself, to overcome certain barriers. Training of all grades of assessors and scheduling the WPBA into the programme are some of the main strategies to overcome problems. Table 3 summarises the strategies used to overcome problems with WPBA [20]. Use of electronic handheld

devices to upload the portfolio can save time and will give the educational supervisor an opportunity to view details of each assessment at any time.

Table 3: Potential strategies for successful implementation of WPBA

Strategy	Criteria	Potential assessment strategy
Watermark	Prolonged engagement	<ul style="list-style-type: none"> • Training of assessors. • Provide information for the assessment. • Incorporate intermittent feedback cycles
	Triangulation	<ul style="list-style-type: none"> • Many assessors, especially different credible groups • Use multiple sources of assessment methods. • Organize a sequential judgement procedure
	Peer debriefing	<ul style="list-style-type: none"> • Organize discussion between assessors for benchmarking and discussion of the process and the results. • Separate the roles of assessor vs coach
	Feedback	<ul style="list-style-type: none"> • Incorporate the learner's views
	Organisational structure	<ul style="list-style-type: none"> • An organising committee to discuss any inconsistencies
Transferability	Sampling	<ul style="list-style-type: none"> • Broad sampling with multiple patients and contexts • Provide more weight to narrative information
Dependability	Documentation Audit	<ul style="list-style-type: none"> • External audit reviews • Give opportunity for student appeal

Modified from JAMA.2002; 287(2):226-235

Conclusion

WPBAs are effective in fostering self-directed and deep learning. Implementation of WPBA needs meticulous planning and training for both the trainer and trainee. Obtaining effective feedback and incorporating it into the summative assessment needs to be devised for individual study programmes.

References

1. Epstein RM. Assessment in medical education. *N Engl J Med.* 2007 Jan 25;2007(356):387-96. <https://doi.org/10.1056/NEJMra054784> PMID:17251535
2. Wass V, Van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. *The lancet.* 2001 Mar 24;357(9260):945-9. [https://doi.org/10.1016/S0140-6736\(00\)04221-5](https://doi.org/10.1016/S0140-6736(00)04221-5)
3. Rethans JJ, Norcini JJ, Baron-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. *Medical education.* 2002 Oct;36(10):901-9. <https://doi.org/10.1046/j.1365-2923.2002.01316.x>

4. Singh T, Modi JN. Workplace-based assessment: A step to promote competency based postgraduate training. *Indian pediatrics*. 2013 Jun 1;50(6):553-9.
<https://doi.org/10.1007/s13312-013-0164-3> PMID:23942397
5. Liu, Charleen. An Introduction to workplace-based assessments. *Gastroenterology and Hepatology Bed to Bench* 2012;5(1):24-26
PMid:24834194 PMCID:PMC4017451
6. Prescott-Clements L, Van Der Vleuten CP, Schuwirth LW, Hurst Y, Rennie JS. Evidence for validity within workplace assessment: The Longitudinal Evaluation of Performance (LEP). *Medical Education*. 2008 May;42(5):488-95.
<https://doi.org/10.1111/j.1365-2923.2007.02965.x> PMID:18298449
7. Singh T, Modi JN. Workplace Based assessment: A step to promote competency based postgraduate training. *Indian pediatr*.2013;50(6);533-559
<https://doi.org/10.1007/s13312-013-0164-3>
8. Nesbitt A, Baird F, Canning H, Griffin A, Sturrock A. Student perception of workplace based assessment. *Clin teach*:2013;10(6);399-404;
<https://doi.org/10.1111/tct.12057>
9. Salaman Yousuf Guraya. Workplace-based assessment; applications and educational impact. *Malays J Med Sci*.Nov-Dec2015.22(6)5-10
10. Wilkinson JR, Crossley JG, Wragg A, Mills P, Cowan G, Wade W. Implementing workplace-based assessment across the medical specialties in the United Kingdom. *Medical education*. 2008 Apr 1;42(4):364-73
<https://www.ncbi.nlm.nih.gov/pubmed/18338989>
11. AW Phillips and AE Jones. The validity and reliability of workplacebased assessments in surgical training. *The Bulletin of the Royal College of Surgeons of England* 2015 97:3, e19-e23 <https://doi.org/10.1308/147363515X14134529301147>
12. Wood L, Hassell A, Whitehouse A, Bullock A, Wall D. A literature review of multi-source feedback systems within and without health services, leading to 10 tips for their successful design. *Medical teacher*. 2006 Jan 1;28(7):e185- 91.
<https://doi.org/10.1080/01421590600834286>
13. Moonen-van Loon JM, Overeem K, Govaerts MJ, Verhoeven BH, van der Vleuten CP, Driessen EW. The reliability of multisource feedback in competency-based assessment programs: the effects of multiple occasions and assessor groups. *Academic Medicine*. 2015 Aug 1;90(8):1093-9.
<https://doi.org/10.1097/ACM.0000000000000763> PMID:25993283
14. Ramsey PG, Carline JD, Blank LL, Wenrich MD. Feasibility of hospital-based use of peer ratings to evaluate the performances of practicing physicians. *Academic medicine: journal of the Association of American Medical Colleges*. 1996 Apr;71(4):364-70. <https://doi.org/10.1097/00001888-199604000-00014>
15. Wood L, Campbell I. 360 degree assessment: encouraging results of a 6 year study. In Annual Meeting of the Association for the Study of Medical Education, Liverpool, UK 2004.

16. Van der Vleuten CP, Schuwirth LW, Scheele F, Driessen EW, Hodges B. The assessment of professional competence: building blocks for theory development. *Best practice & research Clinical obstetrics & gynaecology*. 2010 Dec 1;24(6):703-19. <https://doi.org/10.1016/j.bpobgyn.2010.04.001>
17. Prescott LE, Norcini JJ, McKinlay P, Rennie JS. Facing the challenges of competency-based assessment of postgraduate dental training: Longitudinal Evaluation of Performance (LEP). *Medical education*. 2002 Jan;36(1):92-7. <https://doi.org/10.1046/j.1365-2923.2002.01099.x> PMID:11849528
18. Van Der Vleuten CP, Schuwirth LW. Assessing professional competence: from methods to programmes. *Medical education*. 2005 Mar;39(3):309-17. <https://doi.org/10.1111/j.1365-2929.2005.02094.x> PMID:15733167
19. Al-Kadri HM, Al-Kadi MT, Van Der Vleuten CP. Workplace-based assessment and students' approaches to learning: a qualitative inquiry. *Medical teacher*. 2013 Apr 1;35(sup1):S31-8. <https://doi.org/10.3109/0142159X.2013.765547>
20. Epstein RM, Hundert EM. Defining and assessing professional competence. *Jama*. 2002 Jan 9;287(2):226-35. <https://doi.org/10.1001/jama.287.2.226>
<https://doi.org/10.1001/jama.287.2.226>