

# The Utility of Work-Based assessments in higher general surgical training – a systematic review

Kamal Raj Aryal (✉ [rkamal2000@hotmail.com](mailto:rkamal2000@hotmail.com))

University of Dundee

---

## Research article

**Keywords:** work-based assessment, general surgery, medical education, surgical training

**Posted Date:** October 31st, 2019

**DOI:** <https://doi.org/10.21203/rs.2.16654/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

**Background** In the UK, work-based assessments (WBAs) including procedure based assessments (PBAs), case based discussion (CBDs), Clinical evaluation exercise (CEXes) and direct observation of procedural skills (DOPS) are used in Higher General Surgical Training Programme (HGSTP). This review aims to investigate trainer and trainee's perception of the usefulness of WBAs based on the published literature in HGSTP.

**Methods** Using mesh headings WBA or PBA or DOPS or Mini CEX with their full forms literature search was carried out in December 2018. Seventeen surgical studies were retrieved describing their usefulness. The usefulness was analysed according to van der Vlugten's utility formula, the product of educational impact, validity, reliability, acceptability, cost-effectiveness and feasibility.

**Results** Among 6 studies on PBA, the validity, reliability, acceptability appeared good. The educational impact was positive to the Kirkpatrick level 1 and 2. One study on Mini CEX showed Kirkpatrick level 1 positive satisfaction in trainees and trainers. CBD had positive Kirkpatrick levels 1 and 2 impact and was valid and reliable by trainees in 2 studies. Two studies on DOPS showed good construct validity with positive Kirkpatrick level 1 impact. Based on 6 studies, the use of multiple methods as used in the intercollegiate surgical curriculum project (ISCP) portfolio was more negative in Kirkpatrick level 1 and 2. The recognised problems are lack of time, lack of faculty development and concerns about their validity and reliability.

**Conclusion** Although the individual WBAs appeared useful in study settings, their perceived usefulness declined when used as multiple methods in practice.

## Background

In line with changes in the world, the post-graduate training programme in the UK has undergone several changes—some of these include European Working Directive (EWTD), Modernising Medical Careers (MMC) and Postgraduate Medical Education and Training Board (PMETB).<sup>1</sup> The traditional apprentice-based training model and assessment during post-graduate training laid a strong emphasis on the knowledge component of the syllabus using multiple-choice questions (MCQs), clinical examinations and vivas.<sup>2</sup> There is increasing public demand and scrutiny on the quality of care provided by doctors and their training received.<sup>3</sup> There was a need for the assessment methods to test them in action (at the workplace) as the MMC was born in 2007.<sup>4,5</sup> Workplace-based assessment (WBA) is one of these systems and refers to “the assessment of day-to-day practices undertaken in the working environment”—or, more simply, workplace-based assessment is an “assessment of what doctors do in practice.”<sup>6</sup>

In the UK, after completing medical school the new doctor enters a long training pathway for several years before becoming a consultant. For surgical specialties, after completion of a 2-year foundation year training, they enter a 2-year generic core surgical training programme after which through a competitive

national selection process, they enter into Higher Surgical Training Program (HSTP).<sup>7</sup> The UK HSTP has 10 surgical specialties namely - Cardiothoracic Surgery, General Surgery, Neurosurgery, Oral and Maxillofacial Surgery (OMFS), Otolaryngology (ENT), Paediatric Surgery, Plastic Surgery, Trauma, and Orthopaedic Surgery (T&O), Urology and Vascular Surgery. General Surgery is one of the largest specialties with an intake of approximately 100 trainees each year through the National Selection process.<sup>8</sup> Several institutions are responsible for training in Surgery. The Royal Colleges of Surgeons through the Joint Committee on Surgical Training (JCST) and its ten Specialty Advisory Committees (SACs) e.g. General Surgery SAC for general surgery set up the curriculum standards for General Surgery. Schools of Surgery at Deanery level and Hospital Trusts at the local level run General Medical Council approved training programmes. The curriculum, delivered through Intercollegiate Surgical Curriculum Project (ISCP), lays strong emphasis on specialty knowledge, clinical exposure, technical and operative skills and professional skills and behaviour.

The ISCP has laid a strong emphasis on WBAs. These include Procedure Based Assessments (PBA), Clinical Evaluation Exercise (CEX), Case-Based Discussion (CBD), Direct Observation of Procedural Skills (DOPS) and multi-source feedback (MSF) to achieve skills of 'performance' at the workplace. The main purpose of the WBAs is to help the trainee to learn and develop and provide evidence of progression in attaining clinical competency.<sup>9,10</sup> They are supposed to be used as a source of formative assessment and provide feedback helping trainee's learning and development. The MSF has been used for many years in many specialties and has been shown to help trainees to develop.<sup>11</sup> The areas assessed by MSF are doctor's behaviour and are carried out by all members of the team where the doctor works. Other WBAs need to be carried out by those who are consultants or trainee's supervisors and assess knowledge, skills and attitudes at specific tasks at a professional level. The evidence available supporting the use of PBA, CEX, CBD, DOPS in higher general surgical training programme (HGSTP) particularly at the time of their introduction was lacking.<sup>12</sup>

The workplace for HGST may be different and unique to many other programmes. The HGST sees and treats patients in the ward as inpatients, sees patients in the outpatient clinics, performs operations to treat several conditions and looks after them pre and postoperatively. The skills required to develop for completion of training are different from other specialties particularly operative skills. This is also a transition between core surgery training (basic surgery training) and becoming an independent practitioner as a consultant. This stage of HGSTP has some independence in many areas where the trainee is already competent to perform tasks at workplace while many actions need to be performed under the direct supervision of the consultant particularly surgical procedures. The WBAs used by this grade need to take into consideration to include all types of work they do in their place of work and need to be useful. This review was undertaken to assess evidence relating to trainee and trainer experiences and perception on the usefulness of WBAs particularly PBA, CBD, DOPS and mini CEX in HGSTP since their introduction.

## Methods

## Search strategy

Using mesh headings work-based assessment or workplace-based assessment or procedure based assessments or PBA or direct observation of procedural skills or DOPS or clinical evaluation exercise or CEX or Mini CEX or Case based discussion or CBD, the literature search was carried out in December 2018 using Medline database. Altogether 3440 articles were obtained as titles dating back to 1997. All these titles in Pubmed were screened by KA and only those articles describing different types of WBAs in all specialties were selected which obtained 158 articles. The abstracts of these articles were screened and only those which were in surgical specialties were selected. This narrowed down the articles to only 30. All of these articles were further studied and only those studies with participants describing usefulness were included which yielded 17 studies. All these 17 studies have been included in following tables. Inclusion criteria included any study either quantitative or qualitative describing WBAs in surgical specialties. Endoscopic procedures, gynaecological procedures, anaesthesia, renal, paediatrics, histopathology nursing, medical students, medical specialties have all been excluded from analysis.

Emphasis has been laid in general surgery training literature though WBAs used in other surgical disciplines have been included. There is some mention of WBAs from other specialties to describe history and background.

## History of WBAs in UK

A study of WBAs, carried out in the UK in 2003 and 2004, to evaluate their use in medical specialties showed them to be feasible and made a reliable distinction between doctor's performances.<sup>13</sup> With the introduction of MMC in 2007, they were then introduced in all specialties including surgery and for all grades in training. Since their introduction, they have been refined, developed and modified. For example, WBAs have been renamed to supervised learning events (SLE) and modified to emphasise feedback for foundation trainees.<sup>14</sup> There are some specialty-specific WBAs, for example, PBAs are used in surgical specialties to include performance in actual operative procedure in the operating theatres while others including DOPS, CBD, mini-CEX are common for all specialties.

## Types of WBAs in HGST and definitions (PBA, Mini CEX, CBD, DOPS)

As discussed above MSF has is a very important and useful marker of doctors' behaviour in all specialties including HGST and not a focus of this review. We will explore instead other WBAs whose evidence base is not as strong in HGST. Table 1 shows the definition of 4 types of WBAs used in HGST.

*Table 1 types of WBAs in HGST*

# Educational Theory Underpinning WBAs

## Miller's pyramid

Miller's pyramid is a framework for assessing clinical competence proposed by psychologist George Miller (see Figure 1).<sup>22</sup> In this pyramid from the lowest level to highest in the sequence is knowledge (knows), followed by competence (knows how), performance (shows how), and action (does). Work-based assessments represent methods of assessment at the highest level of the pyramid. With the help of an assessment in 'action', we can know what happens in the workplace rather than artificial testing conditions at lower levels. The trainee's 'action' collected in this way gives us information about trainee's performance in their day to day practice. Other methods of assessment such as MCQs, simulation tests and OSCEs represent the lower levels of the pyramid. These assessments targeting the highest level in the pyramid can be carried out by the assessor observing trainees in the workplace.<sup>22</sup>

## Utility formula

The usefulness or utility of an assessment has been defined as the product of educational impact, validity, reliability, cost-effectiveness and acceptability.<sup>2</sup> Subsequently, practicality (feasibility) has been added to one of the components of the above equation.<sup>23</sup>

## Impact of WBAs (Kirkpatrick)

To evaluate the impact of teaching and learning programmes, Kirkpatrick (1967) suggested a hierarchical model of evaluation pyramid from lower-level to higher-level from 'level 1—'satisfaction', to level 2 - 'learning', to level 3 -'behaviour' to level 4 - 'results' (Figure 2).<sup>24</sup> Educational impact can be assessed by Barr's adaptation of Kirkpatrick's four-level model on slightly different levels: Level 1—'learners' reactions', Level 2a—'attitudes/perceptions', Level 2b—'acquisition of knowledge/skills', Level 3—'change in behaviour', Level 4a—'change in organizational practice', and Level 4b—'benefits to patients/clients'.<sup>25</sup> Two systematic reviews so far have looked at WBAs and found level 1 or level 2 impact similar to participant "satisfaction" and "learning" in the original Kirkpatrick level.<sup>12,26</sup> Most of these WBAs assessed were mini CEX and DOPS in non-surgical context. The highest level in Kirkpatrick's model is whether the society has been able to get the desired results and patient outcomes are better from the doctors in training who have used WBAs. There is no study proving that WBAs have made a positive impact on patient outcomes/results to Kirkpatrick's level 4.

## Gold standard measures applied WBAs (definitions of Validity, Reliability, Cost Effectiveness, Acceptability,

## Feasibility)

*Validity* means the ability of the WBAs to measure the attribute it sets out to measure. Types of validity include face, content, construct and criterion. Face validity means how close the assessment (WBA) in our context is to the real-world situation. Content validity tells us how closely the WBAs reflect what it sets out to measure (i.e. the particular skill involved). We need to make sure all the topics from the syllabus are covered to get good content. Construct validity is defined as “the degree to which a test measures what it claims, or purports, to be measuring”.<sup>27</sup> Criterion validity includes predictive validity meaning how well the WBA predicts future performance and concurrent validity meaning how well the WBAs correlate with other assessment tools thought to be ‘gold standard’.<sup>28</sup>

*Reliability* refers to the ability of the WBAs to produce consistent, reproducible results again and again.<sup>29</sup> Multiple assessors help to achieve this. PMETB (2005)<sup>30</sup> states—“WBAs are not regarded as being reliable enough to be used as a ‘standalone’ assessment of competence. Hence, they are supposed to be used along with traditional methods of assessment in the form of examinations”.

*Cost-effectiveness* - The time involved in carrying out WBAs has cost implications in a busy schedule of clinicians. Investing in resources and higher cost is likely to enhance trainee’s learning and will evaluate the trainee’s skills at higher level in Miller’s pyramid does how.<sup>2,21</sup>

*Acceptability*- The WBAs as an assessment tool should be acceptable to both the trainer and trainee. Some degree of flexibility on both the trainer and trainee may make this more acceptable.<sup>22</sup>

*Feasibility (Practicality)* - The feasibility is determined by the availability of resources including equipment, time, availability of patients and assessors.<sup>31</sup>

## Results - What The Evidence Says About Specific Assessment Tools

### **PBA**

6 publications in surgical specialities were found describing PBAs (Table 2). There are two publications from Sheffield resulting from the same study including 81 trainees and 749 PBAs across 348 operations evaluating PBA published around the same time. The first one assessed validity, reliability and acceptability of PBA.<sup>32</sup> The second one compared user satisfaction, acceptability, reliability and validity of three different methods PBA, Objective Structured Assessment of Technical Skills (OSATS) and Non-technical Skills for Surgeons (NOTSS).<sup>33</sup> There was overall good validity including construct validity and acceptability, and exceptionally high reliability (G exceeding 0.8). There was a positive impact at Kirkpatrick’s level 1 and 2 as most felt that PBA was important in surgical education and would use it again in the future. The authors recommended that Specialties that use OSATS may wish to consider changing the design or switching to PBA.” This was a good study but included many other nonsurgical specialities; there was a dropout rate of 21.7% of participants after consenting due to various reasons.

Hunter, Baird and Reed (2015),<sup>34</sup> conducted a nationwide email survey of 616 orthopaedic trainees to assess perceived educational benefit of PBAs. Fifty-three percent of the trainees thought PBAs were useful for the delivery of feedback and valued the formative component from PBAs more than the summative assessment gained (Kirkpatrick level 1 and 2). They thought focusing on quality rather than the number of assessment events may improve the educational benefit gained by trainees. This study was a largescale national study but included only the trainees' views without any view from trainers in the orthopaedics training scheme.

Another study in 60 orthopaedic trainees and trainers using survey questions showed there was confusion on the purpose and method of using them. The comments made about them were mainly negative in Kirkpatrick level 1.<sup>35</sup>

Shalhoub, Marshall and Ippolito (2017)<sup>36</sup> carried out semi-structured individual interviews with 10 surgical trainees across different specialties and different grades including 2 basic surgical trainees to explore their views on the value of PBA. Trainees did identify the significant value of PBAs when used correctly and as a formative tool. The use of PBAs as a summative tool was thought to have limited their educational value. The study would have benefitted from the inclusion of trainers' perspectives (supporting triangulation of findings) and participants were both higher trainees as well as basic trainees from a range of surgical specialties. The last paper highlighted the ease of feedback.<sup>37</sup>

## Table 2

In summary, from these 6 publications to date on PBA (5 quantitative, 1 qualitative); the validity, reliability, acceptability of PBA appears good. Kirkpatrick's level for educational impact is positive to the level of 1 and 2, particularly when used correctly with adequate planning as formative and feedback tools. However, there is some negative Kirkpatrick feeling due to confusion on methods of using them.<sup>32-37</sup>

## ***Mini CEX***

Only one study could be found at surgical specialty in postgraduate year 2 trainees in general surgery. There was Kirkpatrick level 1 positive satisfaction in trainees and trainers.<sup>38</sup>

Several studies evaluating the mini CEX have been performed including one metanalysis but all these are in non-surgical specialties. They have shown a medium combined effect on construct validity<sup>39</sup> and level 1 to 2 positive effect on Kirkpatrick model. Some negative points have been; they can be anxiety-provoking and trainees use lenient assessors to get better scores.<sup>40-43</sup>

## *Table 3 Mini CEX studies*

## ***CBD***

Only 2 studies have been published describing CBDs in surgical specialities one in basic and higher surgical training context and another one in otolaryngology context. From these studies, it can be seen that CBD has been taken generally positively at Kirkpatrick levels 1 and 2 and is valid and reliable. Neither of these included trainers.<sup>44,45</sup> Table 4

Looking outside surgery, one from medicine<sup>46</sup> and another one from paediatrics<sup>47</sup>, CBD showed good Kirkpatrick level 1 and 2 educational impact. However, this was trainer dependent and several barriers were encountered including lack of time and lack of trainer engagement.

## ***DOPS***

A study conducted by ISCP found very low uptake of DOPS in 2007 and 2008. These forms were then modified and called S-DOPS. The new forms have been shown to have good construct validity and the uptake of DOPS has increased in basic surgical trainees. Also, there was inverse relation with the complexity of the case and their use meaning they are used more often by core trainees rather than by HGST.<sup>21</sup> Another study in Ophthalmology showed an increase in satisfaction (Kirkpatrick level 1) from year 1 to year 3.<sup>48</sup>

Table 5

Outside surgery, a survey returned by 25 of the 27 pre-registration house officers completing the assessments was positive with most (70%) feeling that direct observation helped to improve clinical skills.<sup>50</sup>

## ***Multiple methods (WBAs)***

Rather than looking at individual WBAs in isolation, we should look at multiple WBAs covering all types of work undertaken by the trainee at the workplace.<sup>51</sup> In the surgical context, 6 studies looked at the impact of multiple assessment methods on education and training. As shown in table 6 several problems have been identified including negative Kirkpatrick level 1 on educational impact.<sup>51-56</sup>

Table 6

## ***Annual Review of competence progression (ARCP) process and number of WBAs***

All HGSTs undergo an ARCP each year. There is a need to have a minimum number of WBAs for each trainee to achieve satisfactory progress. While the London deanery has set this number to 80, for the rest of the deaneries minimum number of required WBAs is 40 at present. One study found the most favoured number was 18. The quality rather than number should be the priority when checked at ARCP.

## Purpose of WBAs—formative or summative

ISCP (2013) <sup>16</sup> states, “WBAs are meant to provide constructive feedback and support learning in the first instance. WBA validation should be carried out immediately afterward following a clinical encounter. There should not be pass or fail but part of learning and development. Those carrying these out should have relevant qualifications, experience and appropriate training including constructive feedback”. Then ISCP also states “they can be used as a summative assessment at the end of the training placement by the ES and counts towards the end of placement ARCP”.<sup>16</sup> These statements cause some confusion about their use as formative or summative. The WBAs themselves are not supposed to be treated as a summative test, although WBAs can form part of the portfolio of evidence submitted at the annual review.<sup>57</sup> One of the contributing factors for lack of engagement by trainees in the use of WBAs has been because these are used as a summative tool.<sup>58</sup>

## Discussion

The WBAs have been in use in HGST for more than 10 years. There are not many studies in surgery particularly in Mini CEX, CBD and DOPs. The studies published so far are heterogeneous with different study outcomes which makes it difficult to synthesize the results to carry out a meta-analysis. Among these 17 surgical studies to date, most of them are on PBA. Only 8 studies purely from general surgery have included either trainee or trainer data only.<sup>21,36–38,44,51–54</sup> All have assessed educational impact except one study which showed good construct validity using surgical DOPS (S-DOPS).<sup>21</sup> Among these 8, 3 were on multiple methods <sup>51,52,44</sup> one on CBD <sup>54</sup>, 2 on PBA <sup>36–37</sup> and one on DOPS <sup>21</sup>, one on Mini CEX <sup>38</sup>. Also, basic trainees and higher trainees are included together in 4 of these studies <sup>36,21,51,52</sup>.

Ten of these studies included here though not from purely general surgery—many points are similar and apply to higher general surgery also. The most useful WBA from these papers appears PBA then CBD. The DOPS and mini CEX do not appear that useful in HGST.<sup>54</sup> As is seen DOPS popularity increased in basic surgical trainees after they were modified. Also, there was an inverse relationship with the seniority of the training and their use.<sup>21</sup> In HGST mini CEX where a trainee is observed clinical encounters such as history taking, examination, breaking bad news etc may not be important to be observed due to their independence in carrying out these tasks in day to day practice. This may have reduced their usefulness. Taking consent may be useful though finding time to be observed by a trainer is still a problem.

The number of WBAs is mostly set at 40 at present in most deaneries. One study in orthopaedics suggested 18 was mostly favoured number which is also the number used in the foundation training programme.<sup>34</sup> The number should be that they should be enough numbers to maintain reliability and at the same time making sure they maintain quality. Having 80 in some deanery may decrease their quality being used as tick box exercise. The quality of WBAs which can be checked at ARCP or by an educational supervisor may help trainees’ learning. Their use as a formative method rather than a summative method would increase the value as shown in these studies.

Generally speaking, the perception of WBAs is becoming more positive with time both in trainees and trainers and most believe that these tools if used correctly and formatively are useful. They help to develop specific skills. It was interesting that though WBAs, when considered alone had good utility based on van der Vuten's utility formula evidenced by good validity, reliability, acceptability and positive Kirkpatrick level 1 or 2. The same could not be seen in real scenarios when they were used together in the ISCP portfolio. This may be partly because when they are studied under study conditions may be different from how they are used in real practice.

With ongoing change and desire to improve, there is a need for global professional assessment of the trainee and such an assessment tool called Entrustable Professional Activities (EPA) assessed by multi consultant review (MCR) using capabilities in practice (CIPs) have been suggested for the future<sup>59,60</sup>. These are intended to be introduced in the HGSTP next year. While these new ways of assessments including EPAs and GPCS assessed by CIPS and MCRs should be incorporated for professional and global skills; WBAs should be retained for the development of specific skills. It is important that they are used properly so that trainees get the maximum benefit from them. There is a need to revisit and conduct large scale national and international studies to assess the impact of these tools and improve them helping in the development of general surgical trainees.

## Abbreviations

ARCP Annual Review of Competence Progression

CBDCase Based Discussions

CEX Clinical Evaluation Exercise

CIPS Capabilities in practice

Mini CEX Mini Clinical Evaluation Exercise

CEX-C Clinical Evaluation Exercise—Consent

DOPSDirectly Observed Procedural Skills

ENT Ear Nose and Throat

EPA Entrustable professional activities

GMCGeneral Medical Council

GPC General Professional Capabilities

HGST Higher General Surgical Training

HGSTPHigher General Surgical Training Programme

ISCPIntercollegiate Surgical Curriculum Project

JCSTJoint Committee on Surgical Training

MCQMultiple choice question

MCR Multi consultant report

MMCModernising Medical Careers

MSFMultisource Feedback

NOTSSNon-technical Skills for Surgeons

OCAP Orthopaedic Competence Assessment Project

OMFSOral Maxillo Facial Surgery

OSATS Objective Structured Assessment of Technical Skills and

OSCE Objective structured clinical examinations

PBAProcedure Based Assessments

PMETBPostgraduate Medical Education and Training Board

SACSpecialist Advisory Committee

SLESupervised learning Events

T & OTrauma and Orthopaedics

WBAWork Based Assessments

## **Declarations**

- Ethics approval and consent to participate—not required for this review
- Consent for publication—Not applicable
- Availability of data and material—All data generated or analysed during this study are included in this published article
- Competing interests - None
- Funding - None
- Authors' contributions—KA all

- Acknowledgements—Mandy Moffat, University of Dundee, Department of Medical Education, for looking and advising the manuscript

## References

1. Bannon M. What's happening in postgraduate medical education? *Arch Dis Child*. 2006;91(1):68-70
2. van der Vleuten C. The assessment of professional competence: developments, research and practical implications. *Adv Health Sci Educ*. 1996;1:41–67
3. Norcini J, Talati J. Assessment, surgeon, and society. *Int J Surg*. 2009;7:313–17.
4. Modernising Medical Careers (2010). Medical specialty training (England). Accessed in Dec 2018 from [mmc.nhs.uk](http://mmc.nhs.uk).
5. Scheele F, Teunissen P, Van Luijk S, Heineman E, Fluit L, Mulder H, et al. Introducing competency-based postgraduate medical education in the Netherlands. *Med Teach*. 2008;30:248-53.
6. Swanwick T, Chana N. Workplace-based assessment. *Br J Hosp Med (Lond)*. 2009;70:290-3.
7. General Medical Council (2017). Training pathways: analysis of the transition from the foundation programme to the next stage of training accessed in July 2019 from <https://www.gmc-uk.org/-/media/documents/dc10999-evi-training-pathways-analysis-of-transition-from-foundation-to-next-stage-of-train-74522826.pdf>
8. Patel R. How to prepare for ST3 general surgery application. *BMJ*,2016;352,i352
9. General Medical Council (2010). Academy of Medical Royal Colleges. Workplace based assessment: a guide for implementation
10. General Medical Council (2011). Learning and assessment in the clinical environment: the way forward. General medical council
11. Saedon H, Salleh S, Balakrishnan A, Imray CH, Saedon M. The role of feedback in improving the effectiveness of workplace based assessments: a systematic review. *BMC Med Educ*. 2012;12: 25
12. Miller A, Archer J. Impact of work place based assessment on doctors' education and performance: a systematic review. *BMJ*. 2010;24:341:c5064.
13. Wilkinson JR, Crossley Miller A, Archer J. Impact of work place based assessment on doctors' education and performance: a systematic review. *BMJ*. 2010; 24:341:c5064.
14. Rees CE, Cleland JA, Dennis A, Kelly N, Mattick K, Monoroux LV. Supervised learning events in Foundation Programme: a UK - wide narrative interview study. *BMJ Open*. 2014;. 4,e005980.
15. Pitts D, Rowley DI, Sher JL(2005). Assessment of performance in orthopaedic training. *J Bone Joint Surg Br*, 87(1),187–191.
16. Intercollegiate Surgical Curriculum Project (2013),The intercollegiate curriculum, educating the surgeons of the future, Intercollegiate surgical curriculum overview, accessed in Dec 2018 from [https://www.iscp.ac.uk/surgical/assessment\\_overview.aspx](https://www.iscp.ac.uk/surgical/assessment_overview.aspx)

17. Day SC, Grosso LG, Norcini JJ Jr, Blank LL, Swanson DB, Home MH (1990). Residents' perceptions of evaluation procedures used by then- training program. *J Gen Intern Med.*,5,421-6.
18. Norcini JJ, Blank LL, Arnold GK, et al (1995). The mini-CEX (clinical evaluation exercise): A preliminary investigation. *Ann Intern Med.*,123,795–99.
19. Norcini J, Burch V (2007). Workplace-based assessment as an educational tool: AMEE Guide No 31. *Med Teach*,9,855–71.
20. Bloom BS, Engelhart MD, Furst EJ, Hill WH, Krathwohl DR (1956). Taxonomy of educational objectives:The classification of educational goals. Handbook I: cognitive domain. New York: David McKay Company
21. Matthew A, Beard JD, Bussey M. An evaluation of the use of direct observation of procedural skills in the intercollegiate surgical curriculum programme. *Ann R Coll Surg Engl (Suppl)*.2014; 96: e10–13
22. Miller GE. The assessment of clinical skills/competence/ performance.*Acad Med*. 1990: 65, s63–67.
23. Crossley J, Humphris G, Jolly B. Assessing health professionals. *Medical Education*. 2002;36: 800–804.
24. Kirkpatrick, D. (1967). Evaluation of training. In: Craig, R.L., Bittel, L.R. (eds.). *Training and development handbook*. pp. 87–112. New York: McGraw-Hill
25. Barr H, Freeth D, Hammick M, Koppel I, Reeves S (2000). *Evaluations of interprofessional education*. London: United Kingdom Review of Health and Social Care.
26. Lörwald AC, Lahner FM, Nouns ZM, Berendonk C, Norcini J, Greif R, Huwendiek S (The educational impact of Mini-Clinical Evaluation Exercise (Mini-CEX) and Direct Observation of Procedural Skills (DOPS) and its association with implementation: A systematic review and meta-analysis. 2018; 13(6):e0198009.
27. Brown JD (1996). *Testing language programmes*. Upper Saddle River, NJ:Prentice Hall Regents.
28. Tavakol M, Mohagheghi MA, Dennick R. Assessing the skills of surgical residents using simulation. *J Surg Educ*. 2008: 65, 77–83.
29. Reznick RK . Teaching and testing technical skills. *Am J Surg*.1993 Mar;165(3):358-61.
30. Postgraduate Medical Education and Training Board Workplace Based Assessment Subcommittee (2005). *Workplace based assessment*. Postgraduate Medical Education and Training Board.
31. Ram P, Grol R, Rethans J, Schouten B, van der Vleuten C, Kester A Assessment of general practitioners by video observation of communicative and medical performance in daily practice: issues of validity, reliability and feasibility. *Medical education*. 1999; 33: 447 – 454.
32. 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

33. Beard JD, Marriott J, Purdie J and Crossley J. Assessing the surgical skills of trainees in the operating theatre: a prospective observational study of the methodology. *Health Technology Assessment*. 2011. 15, No. 1.
34. Hunter AR, Baird EJ, Reed MR. Procedure-based assessments in trauma and orthopaedic training– The trainees' perspective. *Med Teach*. 2015; 37(5):444-9.
35. Roushdi I, Tennent D . Current usage patterns of procedure-based assessments in the orthopaedic community. *Bull R Coll Surg Engl*. 2015;97:e1–3.
36. Shalhoub J, Marshall DC, Ippolito K . Perspectives on procedure-based assessments: a thematic analysis of semistructured interviews with 10 UK surgical trainees. *BMJ Open*. 2017 ;24 :7(3).
37. James K, Cross K, Lucarotti ME, Fowler AL, Cook TA. Undertaking procedure-based assessment is feasible in clinical practice. *Ann R Coll Surg Engl* 2009; 91: 110–112
38. 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.
39. Ansari A, Ali SK, Donnon T. The construct and criterion validity of the mini-CEX: a meta-analysis of the published research. *Acad Med*. 2013;88(3):413–20.
  
40. Castanelli DJ, Jowsey T, Chen Y, Weller JM. Perceptions of purpose, value, and process of the mini-Clinical Evaluation Exercise in anaesthesia training. *Can J Anaesth*. 2016; 63(12):1345-1356.
41. Malhotra S, Hatala R, Courneya CA. Internal medicine residents' perceptions of the mini-clinical evaluation exercise. *Med Teach*, 2008;30:41
42. Weller JM, Jolly B, Misur MP, Merry AF, Jones A, Crossley JG, et al.. Miniclinical evaluation exercise in anaesthesia training. *Br J Anaesth*. 2009;102:633-41.
43. Weller JM, Jones A, Merry AF, Jolly B, Saunders D. Investigation of trainee and specialist reactions to the mini-clinical evaluation exercise in anaesthesia: implications for implementation. *Br J Anaesth*, 2009;103:524–30
  
44. Phillips A, Lim J, Madhavan A, Macafee D. Case-based discussions: UK surgical trainee perceptions. *Clin Teach*. 2016;13(3):207-12.
45. Awad Z, Hayden L, Muthuswamy K, Ziprin P, Darzi A, Tolley NS. Utilisation and outcomes of case-based discussion in otolaryngology training. *Clin Otolaryngol*. 2015 Apr;40(2):86-92.
46. Mehta F, Brown J, Shaw NJ. Do trainees value feedback in case-based discussion assessments? *Med Teach*., 2013;35(5):e1166-72.
47. Mohanaruban A, Flanders L, Rees H. Case-based discussion: perceptions of feedback. *Clin Teach*.. 2018; 15(2):126-131.

48. Sethi S<sup>1</sup>, Badyal DK<sup>2</sup>. Clinical procedural skills assessment during internship in ophthalmology. *J Adv Med Educ Prof.* 2019 Apr;7(2):56-61.
  49. 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.
  50. Pilgrim EAM, Kramer AWM, Mokkink HGA, van den Elsen L, Grol RPTM, van der Vleuten CPM. In-training assessment using direct observation of single-patient encounters: A literature review. *Adv Health Sci Educ.*, 2001;16,189–199.
  51. Pereira EA, Dean BJ. British surgeons' experiences of mandatory online workplace-based assessment. *J R Soc Med.* 2009;102(7):287-293.
  52. Pereira EA, Dean BJ. British surgeons' experiences of a mandatory online workplace based assessment portfolio resurveyed three years on. *J Surg Educ.*, 2013;70(1),59-67.
  53. Pentlow A, Field The educational value of work-based assessments: a survey of orthopaedic trainees and their consultant trainers. *Journal of Surgical Simulation*, 2015;2: 60–67.
  
  54. Philips AW, Jones AE. The validity and reliability of work place based assessments in surgical training. *Bulletin of the Royal College of Surgeons of England.* 2015;97: e19-e23.
  55. Eardley I, Bussey M, Woodthorpe A, Munsch C, Beard J (2013). Workplace-based assessment in surgical training: Experiences from the intercollegiate surgical curriculum programme. *ANZ J Surg.*, 83(6),448-453 .
  56. Gaunt A, Patel A, Rusius V, Royle TJ, Markham DH, Pawlikowska T .
- 'Playing the game': How do surgical trainees seek feedback using workplace-based assessment? *Med Educ.* 2017;51(9),953-962.
- 
57. Nassrally MS, Mitchell ES, Bond J. Workplace-based assessments: the emphasis on formative assessment. *Med Teach*, 2012;34: 253.
  58. Swayamprakasam A, Segaran A, Allery L. Work-based assessments: making the transition from participation to engagement. *J R Soc Med Open*, 2014;5 (3).
  59. ten Cate O, Scheele F. Competency-based postgraduate training: can we bridge the gap between theory and clinical practice? *Acad Med.* 2007;82:542–7.
  60. General Medical Council (2017). General professional capabilities framework accessed in Dec 2018 from ([https://www.gmc-uk.org/-/media/documents/generic-professional-capabilities-framework-0817\\_pdf-70417127.pdf](https://www.gmc-uk.org/-/media/documents/generic-professional-capabilities-framework-0817_pdf-70417127.pdf))

## Tables

## Table 1 types of WBAs

<b>PBA</b>	This was originally developed by the Orthopaedic Competence Assessment Project (OCAP) for T & O <sup>15</sup> and subsequently refined by the SAC for surgery for use in all the surgical specialties. The form has 3 principal components (ISCP) – feedback trainer and trainee, a series of competencies within 5 domains (preop, consent, exposure/closure, intraoperative, post-op) and finally a global assessment that is divided into 8 levels of competence. The highest competence rating means the trainee is able to perform the procedure without supervision and deal with complications the standard expected of day 1 consultant in the National Health Service (NHS) following CCT. <sup>16</sup>
<b>Mini CEX</b>	CEX was designed originally by the American Board of Internal Medicine and this took 2 hours to complete. <sup>17</sup> The mini CEX a shortened version was then developed which has been modified to fit in the surgical clinical encounters. <sup>18</sup> These include observing the trainee's interaction with patient in events such as history taking, physical examination, consent (CEX-C), professionalism and communication skills in breaking bad news.
<b>CBD</b>	The CBD was originally called chart-stimulated (CSR) by American Board of Emergency Medicine. <sup>19</sup> A modified version called case-based oral assessment was used by GMC to assess failing doctors using 2 assessors who assessed a doctor including a portfolio of cases managed by the doctor. There was a good representation of daily activity. Finally, it was developed into CBD and is used across all specialties. Currently, this is supposed to be conducted as a structured in-depth discussion of management of patient by the trainee and justifications for their actions. This explores knowledge, judgement and clinical reasoning of that trainee testing a higher level of thinking and promoting deep learning in Bloom's Taxonomy. <sup>20</sup>
<b>DOPS</b>	This was originally developed and evaluated by Royal Colleges of Physicians in a range of basic diagnostic and interventional procedures assessing trainee's technical, operative and professional skills. It has been modified to include range of procedures in the surgical context including operating theatres. <sup>21</sup>

## Table 2 PBA

Author	Study design	Assessed	No of participants	Results	Subject area
Marriot et al,2011 <sup>32</sup>	Quantitative	Validity, reliability and acceptability	81 trainees across all surgical and gynae specialties	Good construct validity	All surgical specialties including Gynae
Beard et al,2011 <sup>33</sup>	Quantitative	User satisfaction, acceptability, reliability and validity	81 trainees across all surgical and gynae specialties	G score >0.8 Positive impact Kirkpatrick's level 1 and 2	All surgical specialties including Gynae
Hunter et al, 2015 <sup>34</sup>	Quantitative national survey	Perceived educational value	616 orthopaedic trainees	Positive Kirkpatrick level 1 and 2 by 53% of respondents.  focus should be quality than number of WBAs completed	Orthopaedic higher surgical raining
Roushdi, 2015 <sup>35</sup>	Quantitative survey	Perceived educational impact	60 orthopaedic trainees	Negative Kirkpatrick level 1 Confusion on purpose and method of using them	Orthopaedic trainees and trainers
Shalhoub,2017 <sup>36</sup>	Qualitative semistructured interviews	Perceived value of PBA	10 surgical trainees	Significant value when correctly used and as formative tool	Surgical training across different specialties and grades
James et al, 2009 <sup>37</sup>	Quantitative observational	Time required to do PBA forms and ease of use PBA	3 trainers 2 trainees 7 pilot PBAs and 15 PBAs	Enabled Focussed feedback  Assessment tool is effective if commitment from trainers and trainees and adequate planning,	Higher surgical trainees

Table 3 mini CEX

Author	Type of study	Assessed	No of participants	Results	Subject area
Joshi,2017 38	quantitative	Mini CEX satisfaction	9 trainers, 16 trainees post graduate year 2 general surgery	Feasible, trainees good to high satisfaction  faculty overall good satisfaction	General surgery

Table 4 CBD

Author	Study design	Assessed	No of participants	Results	Subject area
Philips et al, 2016 <sup>44</sup>	Quantitative survey	Usefulness	42 surgical trainees (21 basic and 21 higher trainees)	Positive and negative themes	Surgical trainees
Awad,2015 <sup>45</sup>	Retrospective analysis of database	CBD validity, reliability and	46 trainees, 1400 WBAs	Reliable and valid tool	ENT core and higher trainees

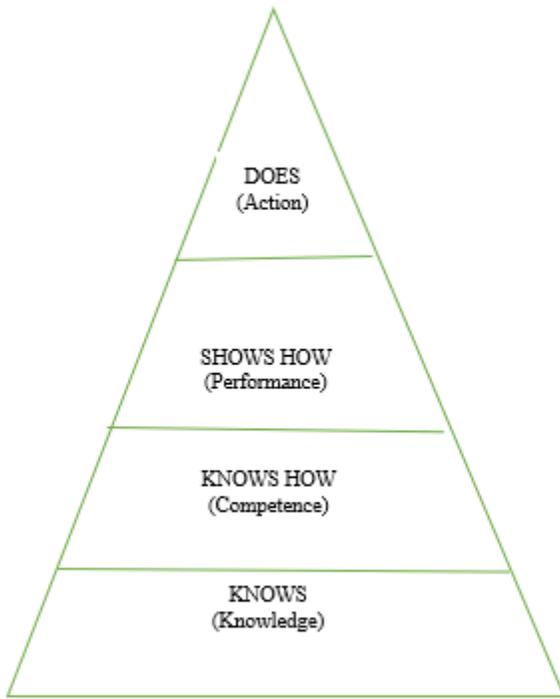
Table 5 DOPS

Author	Study design	Assessed	No of participants	Results	Subject area
Sethi,2019 <sup>48</sup>	quantitative	DOPS	115 ophthalmology interns (trainees)	Increase in satisfaction from 1 <sup>st</sup> to 3 <sup>rd</sup> DOPS	Ophthalmology
Matthew et al, 2014 <sup>21</sup>	quantitative	DOPS uptake  construct validity	Trainees – number not mentioned	Good uptake on modifying the form  New forms good construct validity  Less common in HGST	ISCP portfolio data, surgical trainees of all levels

Table 6 multiple methods of WBAs

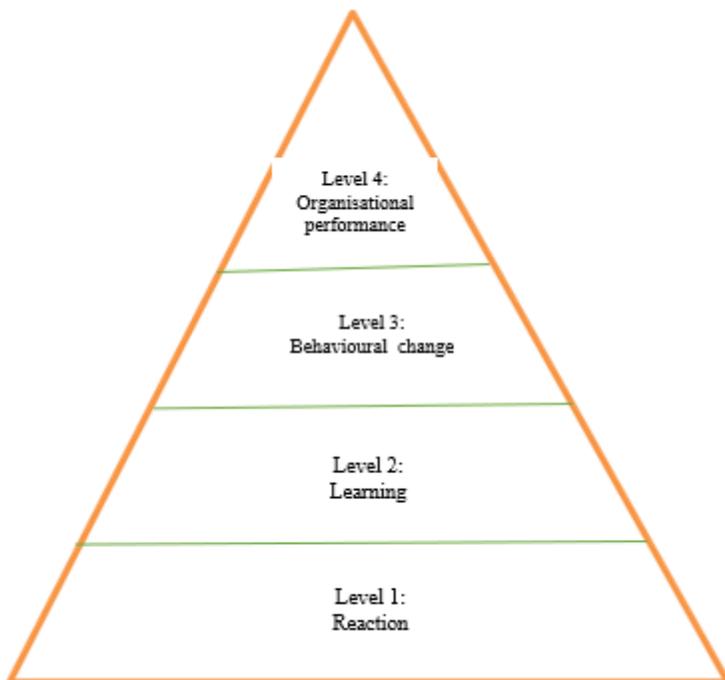
Author	Study design	Assessed	No of participants	Results	Subject area
Pereira et al, 2009 <sup>51</sup>	Survey ISCP portfolio	ISCP portfolio including MSF, mini CEX, DOPS, CBD	539 users (trainees)	90% said it had neutral or negative impact on training  60% said frequently or adversely affected the training (Kirkpatrick level 1)	Surgery
Pereira et al, 2011 <sup>52</sup>	Resurvey of above	ISCP portfolio including MSF, mini CEX, DOPS, CBD	359 users (trainees) across all specialities	Poor by 36% of respondents	Surgery
Pentlow 2015 <sup>53</sup>	Survey	WBAs	61 trainer and 46 trainees	lack of satisfaction	Orthopaedics
Philips et al, 2015 <sup>54</sup>	Survey	PBA, CBD, Mini CEX, DOPS	64 general surgery trainers	WBAs beneficial. PBA and CBD more useful than the rest of them	General surgery
Eardley et al, 2013 <sup>55</sup>	Quantitative and qualitative	Experience of ISCP portfolio 2008-2009 one year	No of trainees  2736 CBD 2701 PBA 2504 CBD 864 DOPS	Paucity of low scores  several problems around faculty development, misuse of assessments, inappropriate timing of assessments, concerns about validity and reliability	Surgery
Gaunt, 2017 <sup>56</sup>	qualitative	WBAs feedback	42 trainees, 12 focus group interviews	Took it test or chance to learn. Those who took chance to learn went on to seek negative feedback	Surgery

## Figures



**Figure 1**

Miller's pyramid



**Figure 2**

Kirkpatrick's model

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [PRISMA2009checklist.doc](#)