

# Direct Observation of Procedural Skills (DOPS) as an Assessment Tool for Surgical Trainees

Muhammad Adil Abbas Khan<sup>1</sup>, Mark Gorman<sup>2</sup>, Lukasz Gwozdziwicz<sup>2</sup>, Zain A. Sobani<sup>3</sup>, Clive Gibson<sup>4</sup>

<sup>1</sup>Specialist Registrar, Department of Plastic and Reconstructive Surgery, Castle Hill Hospital, Hull, United Kingdom

<sup>2</sup>Senior House Officer, Department of Plastic and Reconstructive Surgery, Castle Hill Hospital, Hull, United Kingdom

<sup>3</sup>Student, Medical College, Aga Khan University, Karachi Pakistan

<sup>4</sup>Primary Care Medical Educator and Lecturer, Keele University, Staffordshire, United Kingdom

## BACKGROUND

The introduction of the European Working Time Directive (EWT) resulted in reducing the number of working hours and procedures performed by trainees [1]. Furthermore, the shortening of training programs has caused concern about the competency trainees. In order to overcome this gap and enable doctors to make optimum use of their training, the Modernizing Medical Careers (MMC) introduced competency based workplace assessment tools too [2].

These newer workplace-based assessment tools are structured to provide useful feedback to trainees and trainers. The trainee-led programs encompass the assessment of knowledge, attitudes, behavior and learned skills during day-to-day surgical practice. Direct Observation of Procedural Skills (DOPS) is the most commonly used workplace assessment instrument. DOPS was formally introduced in 2005, when it was piloted by the United Kingdom Foundation Programme [3]. The Intercollegiate Surgical Curriculum Programme (ISCP) has encouraged the use of surgical DOPS, along with other assessment tools, for evaluation of surgical trainees due to its clear and user-friendly format and its applicability to clinical, patient-based situations. Here, we provide an overview of DOPS, its purpose, structure and implementation.

## PURPOSE

Judgment, knowledge base and communication skills form the basis for the growth of a future clinician or surgeon. In the context of surgical techniques and skills, in addition to manual dexterity, the above-mentioned traits form the cornerstone of good patient care. There is good evidence that some surgeons lack such proficiency [4, 5]. A study from the University of Toronto, Canada found that direct observation and evaluation of competence in clinical procedures is not routinely undertaken by educational supervisors [6]. This void in training

evaluation can be filled with the use of surgical DOPS as an assessment instrument.

DOPS as an assessment tool was originally developed by the Royal College of Physicians but its use for junior doctors and trainees has been invigorated in recent years. DOPS is unique in that it tests the trainees ability to apply his knowledge to a particular procedure and provides an assessment of the practical work performed by the trainee on a real patient under the supervision of an experienced surgeon.

DOPS is a highly structured tool, which is most applicable in assessing the mechanistic technicalities of procedural skills. An alternative to DOPS, focusing on assessing history taking and patient interaction skills may potentially be the global ratings scale [7]. A structured form of evaluation is preferable to other crude measures of assessment as structured evaluations result in outcomes that are more reliable and the assessments are more effective [8, 9].

In some training programs structural form of evaluation is replacing other more crude measures of procedural competence with poor validity and reliability such as logbooks and supervisor evaluations [10].

## INSTRUMENT

Several studies have found a lack of rigorous testing of procedural skills [10]. To address this deficiency, DOPS is designed to assess the procedural skills of surgical, medical or general practice trainees at all levels.

The skills assessed range from common simple procedures (e.g. venipuncture at the foundation level) to more advanced surgical skills (e.g. oncologic skin excision and local flap reconstruction under local anesthesia). Importantly, the procedures are performed on actual patients rather than simulations, animal models or cadavers.

The trainees are judged on ten criteria that include:

1. Demonstration of understanding regarding

Conflict of Interest: None declared

This article has been peer reviewed.

Article Submitted on: 23<sup>rd</sup> December 2012

Article Accepted on: 22<sup>nd</sup> May 2013

Funding Sources: None declared

Correspondence to: Dr. Muhammad Adil Abbas Khan

Address: Department of Plastic and Reconstructive Surgery, Castle Hill Hospital, Hull, United Kingdom

Email: [adilaku@hotmail.com](mailto:adilaku@hotmail.com)

Cite this article: Khan MAA, Gorman M, Gwozdziwicz L, Sobani ZA, Gibson C. Direct observation of procedural skills as an assessment tool for surgical trainees. *J Pak Med Stud.* 2013; 3(3):137-140

- the indications, relevant anatomy and technique
2. Informed consent.
  3. Demonstration of appropriate pre-procedure preparation.
  4. Appropriate analgesia or safe sedation.
  5. Technical ability.
  6. Aseptic technique.
  7. Acuity to seek help where required.
  8. Post procedure management.
  9. Communication skills.
  10. Consideration of patient/professionalism

A drawback of DOPS is that it evaluates a specific encounter, which may not be representative of a trainees overall performance, rather than rating based on assessment over a longer period of time and that specific encounter [11].

### BLUEPRINT

According to the General Medical Council (GMC), the body that oversees medical education in the United Kingdom, the content assessment of postgraduate training should be based on all areas of "Good Medical Practice".

DOPS fulfills the standard requirements of the GMC which states that the choice of assessment method should be appropriate to the content and purpose of that element in the curriculum. It more comprehensively covers components like Good Medical Practice, and Relationships with Patients, while partially covering components like Good Clinical Care, Working with Colleagues etc. However, it is found to be lacking insight into components like Probity, Health and Teaching and Training. The use of other assessment tools is advocated to cover aspects of the curriculum that cannot be covered by DOPS [11, 12].

DOPS has the ability to systematically sample the content of the surgical curriculum, appropriate to the stage of training. However, surgical curriculum is diverse and the ISCE recommends taking a multi-dimensional approach to evaluation.

### ABSENCE OF BIAS IN DOPS

The use of Surgical DOPS fortunately is not tainted with biases based on trans-cultural or gender issues. DOPS is an exercise assessing technical skill.

However, limitations in communication skills could provide a hindrance to obtaining informed consent and counseling and communication of

results to patients /relatives. However, the authors believe that these are challenges that can be overcome with practice and with progression along the surgical hierarchy do not affect the performance on surgical DOPS.

The surgical trainee chooses an observer for DOPS as well as the time and type of procedure. This reduces stress levels on trainees by avoiding inflexible deadlines. Also, since each DOP covers a separate procedure from other ones and a different observer is present for procedures, procedural and assessor unfairness is minimized. DOPS have been formulated so that appropriate constructive criticism on procedural skills necessary for optimal quality of clinical healthcare may be fed back to trainees. Hence, trainees may receive professional opinion on any areas where their grades have fallen below 'meeting expectation'. This further minimizes observer bias as each observation must be justified. The trainee then has the flexibility and opportunity to re-organize an assessment of the same procedure to check for improvements. Trainees might want to attempt alternative, optional procedural DOPS. The educational supervisor and final year trainer both eventually view the trainee's e-portfolio with all relevant data, including DOPS carried out [13].

### SELECTION OF COMPETENT EXAMINERS/ ASSESSORS

The ISCP guidelines specify that an "assessor" can be a Consultant, Staff Grade, Specialty Registrar, GP or nurse. There is also a separate tick-box on the DOPS form labeled "Other" for another specialist performing the assessment is performed by a person whose title does not match one of the titles from the printed list. Hence, DOPS assessors can be of various levels but in order for them to carry out an accurate DOPS based assessment they are required to be able to relay useful feedback. Hence the assessors should form both awareness and familiarization of DOPS and all assessment procedures they are involved with. It is of benefit to the assessor to be trained in DOPS assessment as well as in rank ordering, equality diversity and the process of providing constructive feedback [14].

The assessor must also have an insight into the curriculum as well as the level of training of the trainee to be able to "standardize" and assess the candidate keeping in mind what the minimally passing candidate should be able to do. They may then be able to grade the trainee according to the five grades of "not enough evidence / below exp-

-ectation / borderline / meeting expectation / exceeding expectation” which provides a fairer assessment [13].

### **MONITORING OF EXAMINERS PERFORMANCE AND TRAINING**

Every DOPS form, in addition to asking for the title, GMC number, full name and signature of the assessor, has two additional questions for the assessor to answer:

1. The first question asks for the amount of DOPS that the assessor has watched being carried out by a trainee.
2. The second question asks whether the assessor has been trained in using the particular assessment tool. The responses to this question has three tick-boxes:
  - a) Face-to-face training
  - b) Have read guidelines
  - c) Web/CD ROM

These questions are to help in the assessment of the assessor to ensure a fair evaluation of the trainee and also to monitor examiner bias based on inadequate assessment experience or lack of training in DOPS based assessment.

By breaking up “marking” of candidates into five awardable grades of not enough evidence / below expectation / borderline / meeting expectation / exceeding expectation in a fair manner, the Wassock factor and examiner flamboyance is reduced [13].

An exceptionally bright trainee may be additionally exonerated and a challenged trainee can be given suggestions for improvement via the “Strengths / Suggestions” box at the bottom on the DOPS form. These questions are to help in the assessment of the assessor to ensure a fair evaluation of the trainee. An assessor code is necessary for information to be entered into the e-portfolio in case of GP trainees.

### **DEFENSIBLE STANDARD SETTING APPROACH**

To minimize unfairness by “dove vs hawk” assessors, the DOPS form has no marks, percentages or grades e.g. A, B, C etc. For each of the 10 items on the form, the assessor can give the trainee one of five possible grades discussed previously.

In addition to these, a grade is given on the general skill in undertaking any procedure. Following the principle of Anghoff Standard Setting, the grades are based on how likely minimally acceptable or competent candidates

are to perform each item correctly [15]. This concept can also be related to the idea of “Minimum Passing Level”.

A core surgical trainee should, at least, gain a minimum grade of 'meeting expectation' in most sections. Not every criteria is applicable for every DOP – and this being the case, the assessor will chose the 'U/C' option to imply the inability to comment as the behavior was not observed. This results in a fairer assessment [13].

However, the standard setting of this exercise depends to a great extent on the assessor’s training level and knowledge of what is expected for a particular procedure from a trainee at a particular level of training. Hence, lack of assessor training may bias the review of a trainee’s performance. An untrained inexperienced assessor may grade an average trainee very highly, or an experienced untrained assessor may expect too much of a junior trainee and label him/her below borderline.

### **FINAL VERDICT**

There is scanty psychometric data on DOPS perhaps due to the fact that direct observation is carried out informally. However intrinsically, in terms of “competency level” it is seen as a high quality instrument as it tests at the “does” level. Authors have commented on the lack of studies accessing the validity and reliability of DOPS, despite it being fairly widely used to assess competency of surgical trainees [16].

Wilkinson et al’s review in 2003 found no validated methods of procedural assessment in literature [17]. However, despite the lack of evidence on its quality, DOPS certainly has good face validity as it is based on the direct observation of a trainee’s procedural skills in real life clinical environments and with real patients [12]. The construct validity of DOPS is explained in studies that document serial improvement in performance of the same procedure by trainees moving up the surgical hierarchy. There is concern that doctors’ behaviors may be influenced if they know that they are being observed due to anxiety and hence DOPS could become a measure of competence instead of being a tool to assess performance [12].

Despite this criticism, the Royal College of Physicians anticipate that DOPS is a highly valid and reliable instrument, particularly when compared to the previous logbook based system [17]. The concurrent validity of DOPS is limited as there is no gold standard. Similarly the predictive validity of DOPS is limited as it

cannot predict future performance. It is anticipated that there will be several future studies on DOPS as part of the instruction to the Foundation Program in the UK [17], and this is especially warranted regarding the reliability of this assessment tool. The main issue of how many procedures should be observed to achieve adequate reliability and also of determining appropriate checklists and rating scales for different procedures needs to be addressed. Since DOPS are assessments performed by a single observer, the issue of inter-rater reliability does not arise for a single assessment. There may be marked differences in performance by the same trainee when performing DOPS testing different procedures. This can have many reasons and does not impose on the reliability of DOPS.

DOPS as an assessment tool is cost effective since it does not require a special set-up or simulated patients/materials. However, the feasibility of DOPS can be influenced (and limited) by the availability of the patient for a particular procedure and the availability of an assessor who is available at short notice when the patient is available. It is often difficult in a busy out-patient set-up of theatre list to find an assessor with enough time to allocate to this in such a short time frame. Also, both trainee and assessor must make sure that they have allocated a suitable length of time in which to perform a DOP. Assessment alone is found to take around 5 - 15 min followed by feedback that lasts for five minutes. In reality, a number of doctors have found that they require a longer time period within which to undertake intimate examinations, gain informed consent and maintain patient dignity. To complete the process, one inevitably uses up more time when having to enter feedback into e-Portfolios. This may again be a little cumbersome in a busy out-patient or theatre list [13].

However, these problems can be overcome with better organization, regular reviews of clinic/theatre lists in advance to see which patients will be attending, enlisting trainers/clinical supervisors to aid in the search for appropriate cases and liaising with departmental consultants and registrars on a regular basis to ensure their availability.

## REFERENCES

1. Lowry J, Cripps J. Results of the online EWTD trainee survey. *Ann R Coll Surg Engl Suppl* 2005; 87:86-87.
2. Delamothe T. Modernising Medical Careers: final report. *BMJ* 2008; 336:54-5.
3. Beard J, Strachan A, Davies H, Patterson F, Stark P, Ball S, et al. Developing an education and assessment framework for the Foundation Programme. *Med Educ* 2005; 39:841-51.
4. Ha JF, Longnecker N. Doctor-patient communication: a review. *Ochsner J* 2010; 10:38-43.
5. Silverman J KS, Draper J. Skills for communicating with patients: Oxford:Radcliffe Medical Press; 1998.
6. Martin JA, Regehr G, Reznick R, MacRae H, Mumaghan J, Hutchison C, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg* 1997; 84:273-8.
7. Essers G, Kramer A, Andriess B, van Weel C, van der Vleuten C, van Dulmen S. Context factors in general practitioner - patient encounters and their impact on assessing communication skills - an exploratory study. *BMC Fam Pract* 2013; 14:65.
8. 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.
9. Carr S. The Foundation Programme assessment tools: an opportunity to enhance feedback to trainees? *Postgrad Med J* 2006; 82:576-9.
10. Sidhu RS, Grober ED, Musselman LJ, Reznick RK. Assessing competency in surgery: where to begin? *Surgery* 2004; 135:6-20.
11. Cipher. Review of work-based assessment methods. Sydney: Centre for Innovation in Professional Health Education & Research; 2007.
12. Hays RB, Davies HA, Beard JD, Caldon LJ, Farmer EA, Finucane PM, et al. Selecting performance assessment methods for experienced physicians. *Med Educ* 2002; 36:910-7.
13. Denny ML. Pulse Registrar: Taking a look at DOPS (Direct Observation of Procedural Skills). *Pulse* 2007; Nov 21.
14. Cohen SN, Farrant PB, Taibjee SM. Assessing the assessments: U.K. dermatology trainees' views of the workplace assessment tools. *Br J Dermatol* 2009; 161:34-9.
15. Zieky MJ. So much has changed: how the setting of cutscores has evolved since the 1980s. In: Cizek GJ, editor. *Setting Performance Standards*. Mahwah, NJ: Lawrence Erlbaum Associates; 2001. p. 19-52.
16. Mandel LP, Lentz GM, Goff BA. Teaching and evaluating surgical skills. *Obstet Gynecol* 2000; 95:783-5.
17. Wilkinson J, Benjamin A, Wade W. Assessing the performance of doctors in training. *BMJ* 2003; 327:s91-2.