

2023-09

Review of the validity of DOPS as an assessment tool for the procedural skills of surgical trainees

Rela, M

<https://pearl.plymouth.ac.uk/handle/10026.1/22286>

10.1308/rcsann.2022.0052

The Annals of The Royal College of Surgeons of England

Royal College of Surgeons of England

All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Please cite only the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.

Review of the validity of DOPS as an assessment tool for the procedural skills of surgical trainees

Abstract

Introduction

Surgical trainees at all stages are mandated to use workplace-based assessments (WBAs) to gain feedback from their trainers. Direct observation of procedural skills (DOPS) is a tool to assess hands-on surgical skills. This review of the literature seeks to ascertain how valid DOPS are as an assessment tool for the procedural skills of surgical trainees according to the APA (American Psychology Association) validity framework.

Methods

Relevant literature was identified through a structured search of Medline, CINAHL and web of science databases, with further papers included on citation review. Following this, papers meeting the inclusion and exclusion criteria were included in the final review specifying those investigating WBAs including DOPS, assessments in practice rather than simulation and specifically for post-graduate surgical trainees.

Findings

16 papers were included in the final analysis. Extracted data from the returned papers was assessed for evidence of validity within each of the five domains on the APA framework: validity based on consequence, response process, reliability, content, and relationship to other variables. There are studies that show good parameters for reliability and validity for specific DOPS used in endoscopy and otolaryngology. However, there is confusion over the

purpose of DOPS amongst trainers and trainees as to whether they should be used formatively or summatively. Recent changes to the surgical curriculum have sought to address this and further work into the impact of this needs to be done.

Introduction

The trajectory for an aspiring surgeon in the UK after graduating from medical school and completing two foundation years, is two years of core surgical training followed by 5-6 years of higher surgical training in a sub-specialty. The method through which surgical skills are assessed in the early stages of surgical training is of critical importance to enable learning and self-improvement.

Direct observation of procedural skills (DOPS) assesses hands-on surgical skills which are a crucial element of a surgeon's practice. The UK Joint Committee on Higher Surgical Training (JCST) introduced DOPS along with other workplace-based assessments (WBAs) into the surgical curriculum in 2007 on the electronic portfolio ISCP (intercollegiate surgical curriculum programme) (1). An example of the DOPS form is in Table 1. This was secondary to a shift in mentality from surgical training being based on an apprenticeship model (2–4) whereby the more years in practice meant the better surgical skills and the trainees' primary positive feedback would be being told they had a "good pair of hands", to a competence-based approach founded on objective and structured feedback (4).

This critical analysis of the literature seeks to assess the five aspects of validity of DOPS according to the American Psychological Association validity framework (5) (content, internal structure, response process, relationship to other variables and consequence).

Methods

A literature review was conducted in February 2021 using the following search terms: (surgeon OR surgical trainee OR surgical resident OR surgery resident) AND (direct observation of procedural skills OR DOPS) NOT (nursing OR dental OR undergraduate) in the in the following databases: Medline, CINAHL and Web of Science. Further papers were also identified on citation review. A PRISMA flow diagram (6) in Figure 1 illustrates the search process.

The inclusion and exclusion criteria used to identify the articles for review are detailed in Table 2.

This review of the literature was not designed to be exhaustive as we were not seeking to measure the impact of an intervention. Rather, we were seeking to gather evidence for the validity of using DOPS as a training and assessment tool. Validity is the *sin qua non* of assessment design (7) because a measure of validity is essentially a measure of whether an assessment measures what it should measure. Historically this kind of exercise was framed in terms of the single concept of “construct validity”. However, the APA validity framework, as described by Downing and others(7–9), divides construct validity into five validity domains listed in Table 3. This facilitates a much more robust process as it allows for consideration of, and demarcation between, different aspects of validity.

Importantly, the APA validity framework includes the domain of consequence validity, which is concerned with the intended and unintended consequences of the assessment. This is much overlooked in validity frameworks that focus solely on the psychometric properties of a test. Given the high stakes of postgraduate surgical examinations, and the need for assessment to drive learning, it is an important validity domain when considering the use of assessment for postgraduate surgical trainees.

Results and Discussion

Three articles were excluded after initial screening due to being either inaccessible or solely reporting quantitative WBA figures. Table 4 summarises the 16 papers included in the final analysis. 7 papers were related to WBAs in general including DOPS. 13/16 were UK articles and all the studies were from 2010 onwards.

Validity based on consequence looks to ensure the assessment is being used appropriately to drive learning. This topic is explored first as it was the main focus of several of the papers reviewed (10–15). Their primary concern was of the ambiguity of the purpose of DOPS and whether they are an assessment of, or for, learning. The General Medical Council's stance outlines the dual role of WBAs being formative in allowing 'learning through constructive feedback' and summative in 'determining competence levels to permit progression of training or certification' (12). The Chairman of JCST sought to clarify the issue by stating that WBAs should be used as 'feedback tools to direct training' rather than summative

assessments (17). Despite this, Gaunt et al., found that trainers and trainees perceive WBAs as predominantly summative assessments (12). Shalhoub et al's paper focuses on procedure-based assessments (PBAs) which also assess procedural skills but for more complex operations. They postulate that PBAs are being used summatively as they are included in the annual review of competence progression (ARCP) which set a quota for the required number of assessments each year (11). Ali et al's review indicates that the JCST's decision to increase the quota from 20 to 40 further pushes WBAs into being a tick-box exercise (14). Moreover, Nicholas et al., provide evidence for WBAs being used in selection processes for higher surgical training, where applicants are asked to provide numbers of DOPS achieved at each 'grade'(18). The use of PBAs and DOPS as a summative assessment likely negates their use as an effective learning tool (11).

The new surgical curriculum of the ISCP, introduced in August 2021, has removed the overall quota for WBAs including DOPS, except for a few critical skills such as skin closure. Instead, it recommends that trainees agree to complete optional WBAs based on their interests, as agreed with their educational supervisors. Figure 2 from the ISCP website describes the contribution of WBAs to the overall assessment of core surgical trainees (19). This removal of the set quota for DOPS and separation of mandatory and optional DOPS appears to push the focus more towards a formative assessment. However, DOPS are likely to retain a summative element as they will continue to contribute to the assessment of clinical skills as part of a supervisor's summative report, thus contributing to the final outcome of the ARCP. Further research into the impact of the new curriculum over the coming years is required to determine the extent of the change in the function of DOPS in surgical training.

Response process refers to the process in which learners use the assessment. In the context of DOPS, how trainees seek to gain feedback will be analysed. Shalhoub et al., point to the fact that half of trainees were filling out their own DOPS forms on behalf of the assessor (11). Trainees are often being asked to do this by trainers of whom, only 10% in one study, claim they have specific time in their job plan to complete WBAs (20).

Gaunt et al's study highlights a number of factors determining whether a trainee asks a trainer to complete a WBA, including their relationship with the trainer, the trainer's credibility and their engagement with the WBA process (12). This highlights the multi-factorial nature of the feedback process and contradicts other papers which comment on trainees' motivations to 'cherry pick' cases where they have performed well (20) and 'game the system' by choosing assessors who they perceive as being less critical (11).

Internal structure is a measure of the reliability of an assessment tool. Siau et al conducted studies on the use of DOPS in gastroscopy and flexible sigmoidoscopy (3,21). They performed a generalisability analysis which pointed to good reliability. Awad et al., analysed the DOPS forms of otolaryngology trainees for parametrics of reliability, which showed positive results (22).

Many studies suggest that it is a lack of assessor training that impacts the reliability of the assessment. However, Phillips et al., claim that all of the trainers in their study had some form of training in completing WBAs and Gaunt et al., report 87% of trainers having attended the 'Training the Trainers' course (12,20). What is likely more useful is regular training for both assessors and trainees (23) rather than a single course to prove competence in providing feedback. Moreover, Gaunt et al., report 70% of trainers would

value the opportunity of being observed providing feedback, suggesting alternative avenues for assessor training (12).

It is difficult to measure inter-rater reliability of DOPS as these assessments are completed by one trainer for a single event. Hassanpour et al., assess the use of video observation of procedural skills (VOPS) which could provide a useful direction for future. This would blind assessments, reduce implicit bias and give trainers more time to provide feedback (24).

Moreover, multiple assessors could be asked to provide feedback on the same procedure to assess inter-rater reliability.

Another difficulty with reliability is that the performance of a surgeon is usually case-specific (14). For this reason, Memon et al., suggest that encouraging trainees to complete a large number of assessments stimulates greater learning from feedback and suggests that it may improve the reliability of the assessment overall, which is a counter-argument to Ali et al's point (2).

Content validity is concerned with what constructs are being measured and whether they are appropriate to the competencies being tested. Siau et al's studies assess the validity of DOPS for two endoscopy procedures for gastroenterologists, nurses and general surgeons. They provide good evidence for content validity due to the multidisciplinary development of these assessments using an 'iterative process of task deconstruction' (3,21). There is no such evidence for how DOPS for surgical trainees introduced by the JCST were created and validated. The studies by Siau et al., may not be that relevant in the context of this analysis as there are only a few endoscopy procedures, compared to the large number of procedures carried out by surgical trainees, making them easier to create evidence-driven assessments for.

Relationship to other variables assesses whether an assessment tool would provide similar results to another valid tool measuring similar constructs. Awad et al., claim that DOPS is valid as it has the capacity to discriminate between the performance of junior and senior trainees (22). This improvement of scores only occurred in the first three years of training and not beyond, but the authors explained this due to DOPS being used to assess more basic surgical skills. The paper claims that because DOPS can differentiate trainees of increasing experience, it could be a useful means of selecting trainees for specialty training. This begs the question of why candidates who have mastered surgical skills should be selected over those who have training potential and the ability to learn from feedback.

This paper, through a systematic search of the existing literature, has sought to assess the validity of DOPs through the application of the APA validity framework domains. Other frameworks, for example Kane et al's (25) breaks validity down into four inferences: scoring, generalisation, extrapolation and implications. However, the APA framework is more useful for this paper due to the inclusion of content and consequence validity. While the psychometric qualities of a test are important, in clinical environments it is essential that a test achieves what it seeks to achieve, and that educators continually evaluate whether: a) the test is mapped to the constructs that are required within that environment; and b), that the balance between the formative and summative aspects of the test are appropriate for its purpose. As such, the use of the APA framework here is one of the strengths of this paper. A limitation of this paper is the potential bias of the lead author MR, being a surgical trainee and having been affected personally by some of the issues described. However, the

second author of the paper is not a clinician and both authors worked together on the interpretation of the study findings to mitigate any potential bias.

Conclusion

The key points from the literature search according to the five validity domains are summarised in Table 5.

Studies assessing parameters for reliability and content validity of specific DOPS i.e. in endoscopy and otolaryngology have shown good results (3,21,22). These were the only studies evaluating DOPS alone, hence why the search was widened to include all WBAs. Trainees have been seen to 'game the system' by selecting cases they have performed well in as well as asking less critical assessors to complete their DOPS in order to get positive feedback. Other points impacting response process validity include the fact that trainees have been seen to fill out their own feedback and time constraints on the part of trainees and trainers impacts the quality and quantity of feedback.

There is conflict in the literature regarding whether mandating trainees to fill out large numbers of DOPS either improves the reliability of the assessment or reduces its usefulness by converting it into a tick-box exercise.

In a competence-based model of education, the summative function of DOPS is essential to ensure that trainees are reaching the required standards of practice. However, feedback on workplace performance is also essential to drive learning in trainee surgeons. This paper shows that there is a clear tension between the formative and summative function of DOPS in surgical training, compromising the consequence validity of this form of assessment.

Changes to the surgical curriculum in the last year have sought to address this by scrapping the WBA quota and creating space for purely formative WBAs. The findings from this study

suggest that this was a move in the right direction. However, further research into the impact of these changes on the validity of these assessments would be valuable and will help to determine whether this tension has been resolved.

References

1. Intercollegiate Surgical Curriculum Programme [Internet]. [cited 2021 Feb 11]. Available from: <https://www.iscp.ac.uk/>
2. Ali AS, Bussey M, O'Flynn KJ, Eardley I. Quality of feedback using Workplace Based Assessments in urological training. *Br J Med Surg Urol* [Internet]. 2012 Jan 1 [cited 2021 Feb 11];5(1):39–43. Available from: <https://www.sciencedirect.com/science/article/abs/pii/S1875974211001790>
3. Siau K, Crossley J, Dunckley P, Johnson G, Feeney M, Hawkes ND, et al. Direct observation of procedural skills (DOPS) assessment in diagnostic gastroscopy: nationwide evidence of validity and competency development during training. *Surg Endosc* [Internet]. 2020 Jan 25 [cited 2021 Feb 11];34(1):105–14. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/30911922>
4. Memon MA, Brigden D, Subramanya MS, Memon B. Assessing the surgeon's technical skills: analysis of the available tools. *Acad Med* [Internet]. 2010 May [cited 2021 Feb 13];85(5):869–80. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20520044>
5. American Educational Research Association, American Psychological Association NC on M in E. Standards for Educational and Psychological Testing [Internet]. 2014 [cited 2021 Aug 27]. Available from: https://www.testingstandards.net/uploads/7/6/6/4/76643089/standards_2014editio

- n.pdf
6. Moher D, Liberati A, Tetzlaff J, Altman DG, Group TP. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med [Internet]. 2009 Jul 21 [cited 2021 Feb 12];6(7):e1000097. Available from: <https://dx.plos.org/10.1371/journal.pmed.1000097>
 7. Downing SM. Validity: on meaningful interpretation of assessment data. Med Educ [Internet]. 2003 Sep [cited 2021 Dec 3];37(9):830–7. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/14506816>
 8. Messick S. VALIDITY. ETS Res Rep Ser [Internet]. 1987 Dec 1 [cited 2021 Dec 3];1987(2):i–208. Available from: <https://onlinelibrary.wiley.com/doi/10.1002/j.2330-8516.1987.tb00244.x>
 9. Messick S. The psychology of educational assessment. In: Linn R, editor. Educational Measurement 3rd Edition. New York: American Council on Education and Macmillan; 1989.
 10. Mayne A, Wilson L, Kennedy N. The Usefulness of Procedure-Based Assessments in Postgraduate Surgical Training Within the Intercollegiate Surgical Curriculum Programme; A Scoping Review. J Surg Educ [Internet]. 2020 [cited 2021 Feb 10];77:1227–35. Available from: <https://doi.org/10.1016/j.jsurg.2020.03.005>
 11. Shalhoub J, Marshall DC, Ippolito K. Perspectives on procedure-based assessments: a thematic analysis of semistructured interviews with 10 UK surgical trainees. BMJ Open [Internet]. 2017 Mar 24 [cited 2021 Feb 11];7(3):e013417. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28341687>
 12. Gaunt A, Patel A, Royle J, Fallis S, Almond M, Mylvaganam S, et al. What do surgeons and trainees think of WBAs and how do they use them? Bull R Coll Surg Engl

- [Internet]. 2016 Oct 29 [cited 2021 Feb 11];98(9):408–14. Available from:
<https://publishing.rcseng.ac.uk/doi/10.1308/rcsbull.2016.408>
13. 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* [Internet]. 2017 Sep 1 [cited 2021 Feb 11];51(9):953–62. Available from:
<http://doi.wiley.com/10.1111/medu.13380>
 14. Jason M. Ali. Getting Lost in Translation? Workplace based assessments in surgical training. *Surg* [Internet]. 2013 [cited 2021 Feb 11];286–9. Available from:
<https://reader.elsevier.com/reader/sd/pii/S1479666X1300036X?token=C68D27F344EF593F5A0E4D2A38D30EBF479D032C4C58718B508B07E9B8D02851EF2FE12FD53E8474A6F01DF7349F6996>
 15. Aryal KR, Currow C, Downey S, Praseedom R, Seager A. Work-Based Assessments in Higher General Surgical Training Program: A Mixed Methods Study Exploring Trainers' and Trainees' Views and Experiences. *Surg J (New York, NY)* [Internet]. 2020 Jan [cited 2021 Feb 11];6(1):e49–61. Available from:
<http://www.ncbi.nlm.nih.gov/pubmed/32158953>
 16. Workplace Based Assessment: A Guide for Implementation. *Gen Med Counc*. 2010;
 17. Munsch C. British surgeons' experiences of mandatory online workplace-based assessment--reply to authors. *J R Soc Med* [Internet]. 2009 Sep 1 [cited 2021 Feb 11];102(9):357–8. Available from:
<http://journals.sagepub.com/doi/10.1258/jrsm.2009.090278>
 18. Rebecca Nicholas JW. Work-Based Assessments: Do not forget the Formative! *J Surg Educ* [Internet]. 2014 [cited 2021 Feb 11]; Available from:
<https://reader.elsevier.com/reader/sd/pii/S1931720414000440?token=AE85F005B16>

- 277F8D136442CE9D400C9A48DA02F3D856B5EC5B9779D4B46773259D6775CBF8697
CC3EEEDE7E1D56505C
19. ISCP - Core Surgical Training Curriculum [Internet]. 2021 [cited 2022 Mar 7]. Available from: <https://www.iscp.ac.uk/iscp/curriculum/core-surgical-training-curriculum/5-programme-of-assessment/>
 20. Phillips AW, Madhavan A, Bookless LR, Macafee DAL. Surgical Trainers' Experience and Perspectives on Workplace-Based Assessments. *J Surg Educ* [Internet]. 2015 Sep 1 [cited 2021 Feb 11];72(5):979–84. Available from: <https://www.sciencedirect.com/science/article/pii/S193172041500077X?via%3Dihub>
 21. Siau K, Crossley J, Dunckley P, Johnson G, Haycock A, Anderson JT, et al. Training and assessment in flexible sigmoidoscopy: using a novel direct observation of procedural skills (DOPS) assessment tool. *J Gastrointest Liver Dis* [Internet]. 2019 Mar [cited 2021 Feb 12];28(1):33–40. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/30851170>
 22. Awad Z, Hayden L, Muthuswamy K, Ziprin P, Darzi A, Tolley NS. Does direct observation of procedural skills reflect trainee's progress in otolaryngology? *Clin Otolaryngol* [Internet]. 2014 Jun 1 [cited 2021 Feb 12];39(3):169–73. Available from: <http://doi.wiley.com/10.1111/coa.12251>
 23. Rauf S, Aurangzeb W, Abbas S, Sadiq N. WORK PLACE BASED ASSESSMENT IN FOUNDATION YEAR: FOUNDATION UNIVERSITY MEDICAL COLLEGE EXPERIENCE [Internet]. Vol. 23, *J Ayub Med Coll Abbottabad*. 2011 [cited 2021 Feb 12]. Available from: <http://www.ayubmed.edu.pk/JAMC/23-4/Shahid.pdf>
 24. Hassanpour N, Chen R, Baikpour M, Moghimi S. Video observation of procedural skills for assessment of trabeculectomy performed by residents. *J Curr Ophthalmol*

[Internet]. 2016 Jun [cited 2021 Feb 11];28(2):61–4. Available from:

<http://www.ncbi.nlm.nih.gov/pubmed/27331148>

25. Kane M. Validating score interpretations and uses. *Lang Test* [Internet]. 2012 Jan 17

[cited 2021 Oct 14];29(1):3–17. Available from:

<http://journals.sagepub.com/doi/10.1177/0265532211417210>