

The Changing Landscape of Workplace-Based Assessment

Vanessa C. Burch

Honorary Professor, Department of Medicine, Faculty of Health Sciences, University of Cape Town, Rondebosch, Cape Town, 7700, South Africa; vcburch.65@gmail.com

Abstract

Health professions education has undergone radical changes over the past 100 years. This has necessitated a shift away from education programmes largely focused on testing knowledge and skills using predominantly written examinations. There has been a shift towards programmes which are intentionally designed with the end product in mind, a competent practitioner, who is able to provide safe, effective and efficient patient-centred health care. This major paradigm shift has placed an increased emphasis on what health care practitioners do in real clinical practice and the need for assessment processes which provide a holistic overview of trainees' progress towards, and achievement of specified abilities required to undertake independent, unsupervised clinical practice. Such assessment processes, necessarily situated in the workplace, are collectively referred to as Workplace-Based Assessment (WBA). This paper provides an overview of the current state of WBA in health professions education, the challenges that need to be addressed and suggestions about the way forward. It is intended as an introductory text to the topic and a source of useful references to more detailed texts in the literature.

Keywords: Health Professions Education, Medical Education, Postgraduate Training, Workplace-based Assessment

1. Introduction

The training of health care professionals has undergone radical changes over the past 100 years (Carraccio et al., 2002; Carraccio & Englander, 2013). These changes have been fuelled by a recognition the medicine is failing to meet the 21st century health care needs of society (Frenk et al., 2010), public demand for better health care (Jones et al., 2001) and the ongoing problem of medical errors with the associated morbidity, mortality and financial cost (Makary & Daniel, 2016). This has necessitated a shift away from education programmes largely focused on acquiring knowledge and skills predominantly tested using written summative examinations (assessment of learning). There has been a shift favouring programmes which are focused on the progressive development of abilities required to practice medicine. The term Outcomes-Based Education (OBE) (Harden, 1999) is used to describe health professions training programmes which are intentionally designed with the end product in

mind – a competent practitioner who is able to provide safe, effective and efficient patient-centred health care. Over time, it has become clear that the specific competencies expected of health care practitioners, underpinned by appropriate knowledge, skills and professional behavior are best articulated as sets of related competencies specific to the roles health care practitioners fulfill in clinical practice (Holmboe & Harden, 2017). This led to formulation of the term Competency-Based Medical Education (CBME), which is defined as 'an outcomes-based approach to the design, implementation, assessment and evaluation of medical education programs, using an organizing framework of competencies' (Frank et al., 2010a) with the explicit purpose of 'preparing physicians for practice that is fundamentally oriented to graduate outcome abilities and organised around competencies derived from an analysis of societal and patient needs' (Frank et al., 2010b). This paved the way for developing and refining competency frameworks in a number of countries, such as the CanMEDS Competency Framework developed by

the Royal College of Physicians and Surgeons of Canada (Frank & Danoff, 2007), Good Medical Practice developed by the General Medical Council in the UK (2013), and the Outcomes Project of the Accreditation Council for Graduate Medical Education in the USA (Swing, 2007). More recently, it has been proposed that clinicians perform complex clinical activities which require mastery of a set of related competencies. For example, treating a broken hip requires the ability to take a relevant history, perform an appropriate physical examination, order and interpret an X-ray of the hip, and then make and execute a treatment plan such as performing surgery to mechanically 'fix' the hip. While each of these activities is important in their own right, they collectively represent an Entrustable Professional Activity (EPA) defined as: "a unit of professional practice that can be fully entrusted to a trainee, as soon as he or she has demonstrated the necessary competence to execute this activity unsupervised" (Ten Cate, 2005). All these advances in health professions education have placed an increasing emphasis on what health care practitioners do in real clinical practice and the need for assessment processes that document increasing mastery in authentic clinical settings.

This need for assessment methods which can be used to document and track the development of competence in clinical practice has been the major drive towards the development of assessment in the workplace, commonly called Workplace-Based Assessment (WBA) (Norcini, 2005; Swanwick & Chana, 2009). WBA aims to provide a holistic assessment of trainees' progress towards, and achievement of specified abilities required to undertake independent, unsupervised clinical practice. What distinguishes WBA from other assessment modalities is its focus on performance in context, 'what health care practitioners actually do in practice' and the provision of feedback to improve performance, (i.e. assessment for learning) (Black & Wiliam, 1998; Brown, 2005). Given this focus, it is easy to understand why CBME, with its strong emphasis on delivering high-quality patient care (Carraccio et al., 2002; Frank et al., 2010a), has played such a major role in the development of WBA methods and processes (Holmboe et al., 2010; Frank et al., 2017).

Much has been written about WBA, and the literature in this field continues to grow. Currently, most of the published work comes from the UK, Australia, New Zealand, and the USA (Kogan et al., 2009; Miller & Archer, 2010; Saedon et al., 2012; Massie & Ali, 2016). While this lends a specific bias to the existing knowledge and its broader

applicability, particularly in less well-resourced settings, the data are a useful starting point for a discussion of the topic and the role of WBA in health professions education. A further limitation is that the research has been conducted in a range of clinical settings using study participants from a wide scope of practice - from undergraduate students to established practitioners and subspecialists in independent practice. This makes it difficult to generalize findings from one context to another and conclusions may not be generalizable to the entire spectrum of clinical training needs. Furthermore, individual tools have distinct purposes and so the generalizability of study findings should be limited to contexts which match those of the research work and the assessment tools used.

While these broad limitations of the existing literature are important to note when seeking guidance and advice about the implementation of WBA, the published data provide a comprehensive overview of the current WBA landscape and emerging trends. So, the purpose of this paper is threefold: update the reader on the current status of WBA in health professions education; highlight existing challenges described in the literature, and look at possible strategies for realizing the full potential of WBA in health professions education. These are critically important issues for health sciences faculties and specialist certification bodies that are tasked by both the profession and the public with the responsibility of ensuring that practitioners entering clinical practice are competent to provide safe and effective patient-centred health care (Norcini, 2005; Frenk et al., 2010).

2. WBA Tools

A plethora of WBA tools have been developed over the past three decades, and there are many descriptions of the tools used (Davies et al., 2005; Norcini & Burch, 2007; Norcini, 2014). There are at least 55 methods just dedicated to the direct observation and assessment of the clinical skills of medical trainees (Kogan et al., 2009). WBA tools can be classified in a number of ways, but the simplest method describes three categories: single event measurements, global performance measures, and aggregation methods (Van der Vleuten & Verhoeven, 2013). The most frequently used single event measures can be categorised as tools requiring direct observation of clinical activities, including patient consultations, practical/surgical procedures and case discussions based on clinical records (Swanwick & Chana, 2009). Global performance

measures on the other hand, provide a longitudinal overview of performance across a wide range of competencies over a period of time. A few examples of the different types of tools are listed and Table 1 provides a brief description of some of the commonly used WBA tools.

- Direct observation of a patient consultation: mini Clinical Evaluation Exercise (mini-CEX) (Norcini et al., 1995), Clinical Encounter Cards (CEC) (Hatala & Norman, 1999) and Clinical Work Sampling (CWS) (Turnbull et al., 2000)
- Direct observation of practical skills: Directly Observed Procedural Skills (DOPS) (Wragg et al., 2003)
- Direct observation of surgical skills: Objective Structured Assessment of Technical Skills (OSATS) (Martin et al., 1997; Bodle et al., 2008), and Procedure-Based Assessment (PBA) (Isles & Pugsley, 2010; Marriott et al., 2011)
- Patient discussions using case notes: Chart Stimulated Recall (CSR) (Maatsch et al., 1983) and Case-based Discussion (CbD) (Davies et al., 2009; Jyothirmayi, 2012)
- Feedback: also called 360-degree Multi-Source Feedback (MSF) includes the use of tools like the mini-Peer Assessment Tool (PAT) (Archer et al., 2008), team assessment of behaviours (TAB; Whitehouse et al., 2007) and patient satisfaction questionnaires such as the PSQ-18 (Marshall & Hays, 1994).
- Global performance measures: In-Training Evaluation Reports (ITER) (Ginsburg et al., 2013)

Internationally, these tools form part of the 'assessment toolbox' of a growing number of national postgraduate training programmes, such as the UK Foundation Programme (Collins, 2010), the Intercollegiate Surgical Curriculum Programme (Beard & Bussey, 2007) in the UK, the CanMEDS-based residency programme in the Netherlands (Scheele et al., 2008), and the Internal Medicine, Emergency Medicine, and Family Medicine residency programmes in the USA (Swing, 2007). In each of these settings, the tools have been adopted and/or adapted for local needs resulting in several variations of the respective tools in use. Easily accessible examples of some of the data capture forms used in the UK Foundation Programme for postgraduate training are provided in an

overview of WBA tools (Norcini & Burch, 2007). The data derived from the different tools can be aggregated over a period of time to form a longitudinal comprehensive picture of the achievement and development of competence by trainees. Two popular methods of aggregation include portfolios, in which a broad range of assessments and learning activities can be collated, and logbooks which traditionally provide a record of procedures performed and/or events/learning activities attended (Van der Vleuten & Verhoeven, 2013). Portfolios are discussed in more detail later.

This brief overview of WBA tools currently in use has focused on the training of medical doctors. While the examples come from the assessment tool boxes of medical training programmes, the principles of WBA hold true for the training of all health care professionals. A full overview of tools used in the training of nurses and other health care practitioners would require a separate overview of the literature which is beyond the scope of this paper. This brief introduction to the scope of WBA tools sets the stage for a more detailed discussion of the current state of WBA, challenges facing WBA and potential strategies for making better use of WBA in health professions education.

3. WBA and Programmatic Assessment

In health professions education, the need for a broad range of assessment strategies by multiple assessors at multiple points in time is widely accepted (Van der Vleuten & Schuwirth, 2005). However, until fairly recently these assessment processes were often individually tagged onto teaching programmes with the broad purpose of determining 'learning done'; there was little recognition given to the educational role of assessment for learning (Black & Wiliam, 1998; Brown, 2005). As this critical role of assessment has emerged over the past 30 years, it has provoked a call for radical changes in assessment practices (Boud, 1990; Brown, 2005; Van de Vleuten & Schuwirth 2005; Shute, 2008). Two other developments in education have also contributed to the call for change. Social constructivist theories on learning paved the way for a paradigm shift from structure- and process-based to competency-based education (Carraccio et al., 2002). Also, rigorous psychometric evaluation of assessment practices has shown that no single performance

Table 1. Short description of some commonly used WBA tools

WBA tool	Purpose	Assessment process*	Rating of performance
Mini clinical evaluation exercise (mini-CEX)	Assessment of a patient consultation in routine practice	Trainees observed by an assessor during a patient consultation in an inpatient, outpatient or emergency department. Assessment focuses on: history taking, physical examination, communication, clinical judgement, professionalism and organization/efficiency. A typical encounter lasts 15-20 minutes, including feedback.	6-point rating scale: 1-2 below expectations, 3 borderline, 4 meets expectations, 5-6 above expectations. Benchmarked against a Year 1 postgraduate trainee. Space to record suggestions for development and an agreed upon action plan.
Directly observed procedural skills (DOPS)	Assessment of non-operative procedural skills in routine practice	Trainees observed by an assessor while performing a diagnostic/therapeutic procedure on a real patient, e.g. taking a blood sample, administering intravenous medication. Requires about 15 minutes observation time and 5 minutes for feedback.	6-point rating scale anchored like the mini-CEX. Space to record suggestions for development.
Procedure-based assessment (PBA)	Assessment of operative surgical skills in routine practice	Trainees observed by an assessor while performing surgical procedures/ operations on real patients, e.g. performing a Caesarean section.	Assessment form specific to procedure/operation being performed. Comments focus on pre-, intra- and post-operative stages of procedure.
Case-based discussion (CbD)	Assessment of clinical reasoning and patient management in routine practice	A structured discussion between assessors and trainees of challenging clinical cases managed by the trainees. The patient record serves as the basis for the discussion to explore clinical reasoning and a patient care plan. Assessment focuses on: record keeping, clinical assessment (making a diagnosis), selection of investigations and referral, treatment given, follow up and planning of future care, professionalism and overall clinical judgement.	6-point rating scale anchored like the mini-CEX. Space to record suggestions for development and an agreed upon action plan.
Mini peer assessment tool (mini-PAT)	Assessment of professional behaviour and attitudes during routine practice by multiple stakeholders e.g. junior and senior clinicians, nursing staff, allied health professionals	Systematic collection of performance data and feedback for individual trainees using structured questionnaires assessing their ability to provide good clinical care, maintain good medical practice, teach/train colleagues, develop and maintain good relationships with patients, and develop and maintain good working relationships with colleagues.	Each section of the questionnaire has specific questions that are rated using a 6-point rating scale anchored like the mini-CEX. Space to record suggestions for development and an agreed upon action plan.

*Assessor is typically a senior clinician responsible for supervision and training in authentic clinical settings

assessment process is a fool-proof way of determining competence (Van der Vleuten, 1996; Norcini, 2005; Pelgrim et al., 2011; Yeates et al., 2013). Although these developments highlighted the need for multiple assessment methods, the real advance came when assessment was recognised to be a programme design issue (Van der Vleuten & Schuwirth, 2005). This led educators to argue for the specific design of programmes of assessment – a

purpose-driven process of selecting and implementing multiple complementary assessment methods as part of the strategic design of a learning programme (Schuwirth & van der Vleuten, 2011). The intention of such programmes is to monitor and document trainees' progression towards achieving specific capabilities, variably referred to as competencies, milestones, or entrustable professional activities (Englander, et al., 2017), thereby

optimising both the learning and decision-making functions of assessment (van der Vleuten, et al., 2012).

The broad range of assessment tools that are included in programmatic assessment are usually categorized using Miller's classic pyramid describing increasing levels of clinical competence (Miller, 1990). The four-tier pyramid includes two lower levels called 'Knows' and 'Knows how' which refer to the theoretical knowledge that is required for clinical competence. The two upper levels of the pyramid, called 'Shows' and 'Does' (in ascending order), describe the practical application of knowledge in increasingly authentic clinical settings. WBA focuses on determining competence in the workplace which addresses the 'Does' level at the apex of the pyramid (Lockyer et al, 2017).

From the perspective of CBME, it is easy to understand that WBA should be an integral part of a programmatic approach to assessment rather than a 'stand-alone' assessment process, which, like any other assessment modality is subject to limitations. Following the launch and subsequent implementation of competency frameworks previously mentioned, there are now multiple examples of CBME training programmes in which a programmatic approach to assessment includes WBA activities; some examples include postgraduate training programmes in Denmark (Ringsted et al., 2003), the UK (Wilkinson et al., 2008; McKee, 2008), the Netherlands (Scheele et al., 2008), Australia and New Zealand (Jurd et al., 2015) the USA (Batalden et al., 2002) and South Africa (Jenkins et al., 2013).

4. Purpose(s) of WBA

Although WBA is increasingly forming part of undergraduate and postgraduate training programmes, the debate continues about whether WBA should be used for formative assessment and feedback (assessment for learning), summative assessment and high-stakes decision-making (assessment of learning), or both. This question is tightly linked to both the assessment needs of competency-based medical education and the discussion about a programmatic approach to assessment. CBME requires a programme of multiple assessment strategies to paint a comprehensive picture of the emerging competencies of trainees in the workplace (Holmboe et al., 2010; Harris et al., 2017). This approach to education is dependent upon high-quality feedback based on multiple encounters with

multiple observers in the workplace to develop a competent practitioner about whom a collective summative judgement can eventually be made based on a rich pool of data collated over the entire period of training (Lockyer et al., 2017). Prior to the final decision the data should not be reduced to a series of 'mini examinations', (i.e. a series of pass-fail decisions) or a set of rankings (Schuwirth & Ash, 2013). This strong focus on assessment for learning favours the use of WBA to inform the development of competencies and a collective decision-making process regarding the achievement of competence for certification and licencing for independent practice.

While the formative intentions of WBA tools (assessment for learning) have been clearly articulated in the literature (Hauer, 2000; Norcini, et al., 2003; Norcini & Burch, 2007), the use of tools with numeric rating scales has provided the health professions education community with the opportunity to determine the psychometric robustness of WBA tools and consider their use for summative purposes. A systematic review of tools for the direct observation and assessment of clinical skills in medical trainees has highlighted the large body of literature exploring the psychometric properties of performance-based assessment instruments (Kogan et al., 2009). Two important findings have emerged from this 'psychometric discourse' (Hodges, 2013). First, the psychometric adequacy of single assessment events is limited (Kogan et al., 2009; Hodges, 2013), and second, the variability of scores awarded by examiners observing the same examinee may not be a simple 'cognitive' problem that can be resolved by intensive training to 'standardise' their behaviour (Crossley & Jolly, 2012; Yeates et al., 2013; Kogan et al., 2014; Sebok-Syer et al., 2017). It appears that this variability may reflect real differences in different aspects of a trainee's performance that are of varying significance to the individual examiners assessing the trainee. The importance examiners attach to specific aspects of the trainee's performance is idiosyncratic and influenced by a range of contextual factors including examiners' personal clinical experience. This issue is discussed in more detail in a later section of the paper. Collectively, these findings have restricted the use of individual WBA tools to make summative assessment decisions regarding registration and certification to practice medicine (Norcini, 2014).

While individual tools are not used for summative purposes, the underlying concept of programmatic assessment is that the collation of sufficient low-stakes

formative events may be used to inform a summative decision about progress or certification at the end of a period of training (Van der Vleuten, et al., 2012). As previously mentioned, this approach to assessment has been adopted for postgraduate training in a number of countries (Ringsted et al., 2003; Wilkinson et al., 2008; McKee, 2008; Scheele et al., 2008; Jurd et al., 2015; Jenkins et al., 2013). The assumed robustness of these summative decisions is based on the predicted reliability coefficients of a specified minimum number of events for each of the instruments included in the programme (Wilkinson et al., 2008). The composite reliability of these types of assessment packages has been confirmed (Moonen-van Loon, et al., 2013). While this finding supports the use of WBA to inform summative decisions in programmatic assessment, a note of caution needs to be sounded about using WBA tools for both summative and formative purposes.

There is an expanding body of literature highlighting the issues which emerge when WBA takes on a summative role. Trainers start to view all the assessment processes as high-stakes summative events, even if they initially have an explicit formative intention (Bok et al., 2013). Furthermore, faculty are conflicted about the dual roles of assessor and mentor/teacher (Govaerts et al., 2007); raise concerns about the effect of personal relationships on grades awarded (McKavanagh et al., 2012) and are influenced by trainees who are upset about their perceived performance (Cleland et al., 2008; Dudek et al., 2005; Rees et al., 2009; Watling et al., 2010; Dennis et al., 2018). Similarly, trainees feel confused about the purpose(s) of WBA and question the educational value of WBA (Basu et al., 2013; Cohen et al., 2009; Sabey & Harris, 2011; Tailor et al., 2014; Lörwald, et al., 2018). They feel that these assessment processes become summative ‘tick box exercises’ (Bindal et al., 2011) and many consider assessment processes which document numeric scores to be ultimately summative, even if their formative intention has been explicitly articulated (Bok et al., 2013).

One approach to dealing with this situation is to ‘rename’ or ‘rebrand’ assessment processes to highlight the difference between their formative and summative uses in programmatic assessment. For example, the General Medical Council in the UK has recommended that WBA activities with a focus on feedback be termed “supervised learning events” (SLEs) and WBA with summative consequences be called “assessment of performance” (AoP) (General Medical Council, 2011). Royal Colleges responsible for postgraduate training in the UK

have)(or are considering) adopting this approach (Parry-Smith, et al., 2014; O’Leary et al., 16). While early data suggest that this approach may be a good idea, (i.e. SLEs do focus on learning) (Cho, et al., 2014; Rees et al., 2014), there are still difficulties ensuring clarity of purpose (Rees et al., 2014) and direct observation with feedback, the key purpose of SLEs (Cho et al., 2014). This highlights the need for aggressive ‘marketing’ when significant changes like ‘rebranding’ of tools are implemented. It is clear that there is still a critical need for intensive education of users, both trainers and trainees, before introducing advances intended to improve the uptake of WBA (Massie & Ali, 2016).

This discussion has highlighted the use of multiple low-stakes ‘assessment’ events to promote learning and inform summative decisions, without each of the individual events serving as a ‘mini-examination’. This concept still needs to be embedded in modern assessment thinking before it is likely to gain significant further traction.

5. Feedback and WBA

The critical role of feedback in WBA has been described (Norcini & Burch, 2007). Furthermore, there is now also evidence that high-quality feedback from a credible source can change clinical performance (Veloski et al., 2006; Watling et al., 2012). Recent reviews of user perceptions of WBA have highlighted that trainees regard feedback as the most valuable aspect of WBA (Miller & Archer, 2010; Sabey & Harris, 2011; Massie & Ali, 2016). Unfortunately, however, there is substantial evidence that teaching faculty are inadequately prepared for this task (Holmboe et al., 2011) and concerns about the type (Hrisos et al., 2008; Castanelli & Kitto, 2011; Jyothirmayi, 2012) and quality of feedback received by trainees in the workplace (Wilkinson et al., 2008; Cohen et al., 2009; Bindal et al., 2011; Sabey & Harris, 2011) continue to be published. Feedback lacks specificity (Cho et al., 2014); tends to focus on what went well rather than what needs attention (Bok et al., 2013); is often delayed which further jeopardizes its validity (Basu et al., 2013; Bindal et al., 2013; Cho et al., 2014; Tailor et al., 2014), and frequently lacks ‘actionable’ learning plans (Cohen et al., 2009). Importantly, feedback is only considered useful if it is perceived to be accurate and credible; negative or inaccurate feedback does not impact on practice improvement (Sargeant et al., 2005). Faculty training focusing on feedback which is informative, behaviour- and task-specific, based on direct observation and timely

(Hattie & Timperley, 2007; Ramani & Krakov, 2012) should be the first priority of faculty development programmes (Holmboe et al., 2011).

6. Psychometric Challenges and a Qualitative Discourse in WBA

Despite the psychometric limitations and challenges of assessment using real patients in the workplace, the call for authentic assessment continues because there is no adequate alternative for determining competence in real clinical practice (Kogan et al., 2009). As mentioned, much work has been done to try and optimize the psychometric robustness of these assessment events. The most commonly used strategies, with varying levels of success, include: multiple sampling by multiple assessors in a range of authentic settings (Kogan et al., 2009); rater training (Holmboe et al., 2004; Cook et al., 2009b); improving the clinical relevance of rating scales (Cook & Beckham, 2009c; Crossley & Jolly, 2012); and, expanding the assessor pool to include other health care professionals and patients (Saedon et al., 2012).

While these approaches to performance assessment have largely anchored WBA in the world of quantitative assessment, there is an emerging discourse calling for assessment practices which look beyond the psychometric paradigm (Schuwirth & van der Vleuten, 2006; Hodges, 2013) because traditional assessment processes fail to recognize 'key issues in the mechanics of the assessment process' (Govaerts et al., 2007). It is widely recognised that performance assessment is a decision-making process in which rater judgements and trainee performance are influenced by many contextual factors. Therefore, learning and competence, as inferred from performance and interpretation of performance, can only be understood 'in situ' (Govaerts & van der Vleuten, 2013). It is argued that the variability of performance data, as a consequence of rater and trainee interactions with contextual factors, is not just 'measurement error' but rather a source of information about true performance variability (Schuwirth & van der Vleuten, 2006). This perspective requires a radical shift to a 'constructivist, social-psychological approach to performance assessment' (Govaerts et al., 2007). Such an approach to assessment requires an increased focus on the context of assessment rather than just the behaviour of raters and rating scales (Govaerts et al., 2007). It also demands the use of expert-derived narrative text describing performance to guide the development of learners'

competence and make high-stakes decisions about the adequacy of demonstrated competence (Govaerts & van der Vleuten, 2013). The use of narrative text to assess competence should not be confused with the use of narrative text to provide feedback to trainees after a formative assessment event (Sebok-Syer et al., 2017).

While the use of narrative text as part of an assessment process is gaining ground, for example, in UK postgraduate training programmes (General Medical Council, 2011; Cho et al., 2014), concerns about the 'robustness' of qualitative data for high-stakes decision-making have been previously expressed in the literature (Roberts et al., 2002). Proponents of this approach to assessment argue that the criteria used to judge the credibility and dependability of qualitative research data (O'Brien, et al., 2014) can also be used to evaluate qualitative assessment data (Driessen et al., 2005). These criteria, which focus on the trustworthiness and authenticity of the assessment process, are based on a constructivist model of assessment (Guba & Lincoln, 1989) and are well aligned with the assumptions that underpin the traditional quantitative concepts of validity, reliability and objectivity. In this model, the trustworthiness of the assessment information is determined by its credibility (internal validity) and dependability (reliability) (Lincoln & Guba, 1985); a number of methodological strategies have been suggested to ensure both (Morse, 2015). While this is certainly a coherent approach to dealing with psychometric scepticism, there is much work that needs to be done to facilitate widespread implementation of this strategy. One of the nagging concerns about narrative comments as a source of information about in-training performance is the use of vague and indirect language to describe trainee performance in order to 'save face' for both raters and trainees (Ginsburg et al., 2016). The mandate to prioritise the optimal use of narrative comments in WPA is clear.

7. Assessor Variability

There is ample evidence in the literature of assessor variability, which is perceived to be one of the major limitations of performance assessment (Norcini et al., 1997; Boulet et al., 2002; Durning, et al., 2002; Margolis et al., 2006; Wilkinson et al., 2008; Weller et al., 2009a; De Lima et al., 2011; De Lima et al., 2013; McGill et al., 2013). There are many assessor-dependent factors which are known to influence performance assessment and to date the main focus has been on improving the quality of rating scales

and training to improve assessor reliability or 'objectivity'. Two randomised controlled studies have shown the limited impact of assessor training (Holmboe et al., 2004; Cook et al., 2009b). More recently, however, there has been an increasing recognition that assessor variability requires further careful consideration because it may not simply be an issue of assessor 'error' (Kogan et al., 2011; Crossley & Jolly, 2012; Yeates et al., 2013; Govaerts et al., 2013; Kogan et al., 2014; Sebok-Syer et al., 2017). It has been shown that assessors observing the same assessment event pay attention to different aspects of performance, have different interpretations of competence based on their own experience, and generate global impressions of performance which are uniquely expressed and then 'translated' into domain-specific rating scale scores (information integration) (Kogan et al., 2011; Yeates et al., 2013). Furthermore, the complexity of the relationships between checklist scores, task ratings, global ratings and written comments highlight the multiple nuances of performance assessment that contribute to decision-making (Sebok-Syer et al., 2017). These studies flag the need for more work to better understand the non-error sources of assessor variability and their meaning.

8. Rating Scales

The wide range of rating scales used in WBA bears testimony to ongoing attempts to improve the quality of performance-based assessment instruments (Kogan et al., 2009). This has resulted in an ever expanding pool of new tools which have mostly been evaluated in small/pilot studies with little or no data supporting their psychometric rigour (Lurie et al., 2009). While 'assessor quality' and the interpretation of 'assessor variability' may be considered more important than 'fixing the forms', it is worth mentioning the improvements to rating scales that have been suggested in the literature. A number of strategies have met with little success, for example, range restriction (Cook & Beckham, 2009c) and the addition of behavioural anchors (Donato et al., 2008). Currently, there are three strategies receiving ongoing attention: improving the meaning of rating scale anchor points for clinician-educators who use the tools; removing numeric rating scales from formative WBA tools; and, 'reorienting' measurement scales to focus on the quality of patient care provided rather than just the level of competence demon-

strated by the trainee. It has been argued that the use of rating scales which are not anchored in the clinical reality of assessors (i.e. practicing clinicians), is one of the major limitations of existing tools (Crossley & Jolly, 2012). For example, many rating scales use abstract anchors such as 'satisfactory, good and excellent' or 'below, at, or above the expected level of competence' which are not anchored to an observable point of reference (Crossley & Jolly, 2012). Such anchors are problematic because they do not resonate with the 'language' of clinicians (Crossley & Jolly, 2012). Furthermore, it is recognised that clinician-educators do not share a universal understanding of the term 'competent' (Kogan et al., 2011; Crossley & Jolly, 2012; Yeates et al., 2013).

The ongoing perception that rating scales imply a summative purpose for an assessment event has led to the revision of WBA tools in postgraduate training in the UK. Rating scales have been removed from formative assessment activities, now called supervised learning events, and have been replaced by narrative text boxes (Parry-Smith, et al., 2014). This has shifted the focus from a 'score' to a description of observed performance accompanied by feedback about ongoing training needs. This increased focus on feedback is better aligned with the formative intentions of WBA and a more nuanced qualitative approach to assessment for learning.

Finally, there is an emerging concept about shifting the frame of reference of assessment from the trainee to the patient – making assessment events 'patient-centred' rather than 'trainee-centred' (Kogan et al., 2014). The practical implication of this re-orientation is that the 'midpoint' of any measurement instrument should 'equate to the trainee's ability to provide safe, effective, patient-centred care in independent practice' (Kogan et al., 2014). In this paradigm, the care that the patient receives is central to the assessment of competence, and the real measure thereof is the extent to which the clinician-educator participating in the assessment process has to contribute to the patient encounter to secure good care for the patient. Essentially this approach calls for a shift in focus from 'how much the trainee did' to 'how much the trainer needed to do' to ensure that the patient was well treated. Future studies are needed to evaluate the implementation of this concept and its impact on patient care. This may represent an important step in the process of gathering evidence about the impact of WBA on patient outcomes, a glaring gap in the WBA literature.

9. Assessor Training

Assessment ability is acquired and not innate (Lockyer et al., 2017). It follows therefore, that assessor training is an indispensable part of any assessment programme. Assessors require knowledge of the competencies to be assessed (Ponnamperuma, 2013) and an understanding of the correct administration of the assessment process (observation and recording tasks intrinsic to the instrument) (Kogan & Holmboe, 2013). There is good evidence that this level of basic training has not been universally achieved in WBA programmes and the need to address this type of training is not a matter for debate (Massie & Ali, 2016). As has already been discussed, assessor training aimed at reducing score variance has met with little success (Holmboe et al., 2004; Cook et al., 2009b). So, more recently, attempts to improve assessment have shifted to other aspects of the process that may yield more positive results.

First and foremost, is the need for better feedback. Numerous studies have called for assessor training in the delivery of good feedback and some authors are suggesting that certification should be a requirement in the future if a serious attempt at addressing this concern is to be made (Norcini & Burch, 2007). Studies showing that assessor training improves the quality of feedback are emerging in the literature (Pelgrim et al., 2012; Basu et al., 2013). This is a matter which requires ongoing work.

Authors have also put out a call to 'leverage the rater assessment problem' by providing faculty development opportunities for clinical supervisors to hone their clinical skills – the same skills they are assessing (Kogan et al., 2014). This is based on the problematic observation that clinicians who are known to have variable levels of clinical skills expertise (Ramsey et al., 1993; Paauw et al., 1995; Braddock et al., 1997; Vukanovic-Criley et al., 2006; Mitka, 2008) often benchmark trainees against themselves (Kogan et al., 2010). Assessment practices are more likely to have a positive impact on patient care if the clinical skills expertise of training faculty is improved.

10. Training and Development Benefits of WBA

Two of the main anticipated outcomes of WBA is that it should have a positive educational impact on trainees and improve their performance in the workplace. Two reviews have been conducted to address this topic. Miller and Archer (2010) conducted a systematic review

of 16 studies in postgraduate trainees to determine the educational and performance impact of MSF (eight studies), mini-CEX (four studies), DOPS (one study) and a combination of methods (three studies). Fifteen of the studies were non-comparative (descriptive or observational) and of variable quality (Buckley et al., 2009). They found that the best evidence for positive changes in response to WBA came from the studies focusing on MSF. Studies using self-reported data found that MSF was of educational value (Murphy et al., 2009), could lead to changes in attitude (Lockyer et al., 2003; Burford et al., 2010) and even changes in behaviour (Sargeant et al., 2003, 2005; Brinkman et al., 2007; Violato et al., 2008). They noted that changes in performance were dependent on credible, accurate feedback (Sargeant et al., 2005) and coaching to help trainees identify specific strengths and weaknesses (Brinkman et al., 2007). Studies looking at the mini-CEX also found self-reported evidence of a positive educational impact (Malhotra et al., 2008; Nair et al., 2008; Weller et al., 2009b) and improvements in the quality and frequency of feedback (Weller et al., 2009a). In the DOPS study, trainees reported that direct observation improved their clinical skills (Morris et al., 2006). In this review, the findings of studies evaluating the impact of programmes using multiple WBA methods were heterogeneous (Ryland et al., 2006; Wilkinson et al., 2008; Pereira et al., 2009).

A more recent review by Massie and Ali (2016) also explored the contribution of WBA to training and professional development from the perspective of both trainers and trainees. While dental trainees (Grieveson et al., 2011; Kirton et al., 2013) and some core trainees in the UK Foundation Programme appreciated the training value of WBA (Dean & Duggleby, 2013), others did not (McKavanagh et al., 2012; Basu et al., 2013; Rees et al., 2014; Tailor et al., 2014). Postgraduate trainees in surgery (Pereira & Dean, 2009), dermatology (Cohen et al., 2009), general practice (Sabey & Harris, 2011), anaesthesiology (Bindal et al., 2013), and mixed cohorts of specialist trainees (Cho et al., 2014) expressed negative opinions about the educational effects of WBA. Other studies, however, described trainees who expressed support for the use of WPA tools, including trainees in obstetrics and gynaecology (Bodle et al., 2008), histopathology (Finall, 2012), paediatrics (Mehta et al., 2013), psychiatry (Simmons, 2013), ophthalmology (Tsagkatakaki & Choudhary, 2013), anaesthesiology (Weller et al., 2009a), and a mixed cohort of specialist trainees (Fokkema et al., 2013).

A few studies included in the review by Massie and Ali (2016) explored trainers' perceptions of the educational value of WPA. Their findings were generally inconclusive since most of the studies included only a few participants and the results were heterogeneous. Some trainers felt that the time could have been better spent on other training activities (Bodde et al., 2008; Powell et al., 2014) while others recognised some aspects of training benefit (Bodde et al., 2008; General Medical Council, 2018).

Although a broad overview of the prevailing opinions of trainers and trainees is informative, understanding the reasons for engagement or non-engagement with WPA would be more helpful in determining the future use of these assessment tools.

11. WBA to Improve Patient Care

There is a wealth of literature describing the mortality, morbidity and financial costs of errors in medical practice (Makary & Daniel, 2016) and the public demand for safe, effective patient-centred care (Kogan et al., 2014) is widely acknowledged. This underpins the need for a process which directly links clinical training activities to improving patient care. While the intention of all clinical training is to produce health care practitioners capable of delivering good patient care, the direct and explicit link between process and outcome is lacking. For example, it is known that competence does not reliably predict clinical performance in the workplace (Rethans et al., 2002). A paper has highlighted this fundamental oversight and argues powerfully for a strategy to address the problem (Kogan et al., 2014). The authors make a case for changing the frame of reference of WBA from the trainee to the patient. Making the patient the centre of attention redirects the focus of the assessment process towards the ultimate beneficiary of the encounter. Indeed, if trainees secure educational benefits during clinical assessment activities but patients do not receive proper care, the exercise is futile. This approach to assessment has significant implications which are discussed later.

12. Feasibility and User Acceptance of WBA

It is widely recognised that user acceptance of, and engagement in health professions education innovations are critical to successful and ongoing implementation

(Massie & Ali, 2016). Unfortunately this is the Achilles heel of WBA which needs to be addressed if authentic assessment is going to contribute to improved patient care. So, any discussion of WBA should focus on identifying the key challenges that need attention rather than building a case for abandoning WBA because there is widespread 'cynicism' (Academy of Medical Royal Colleges, 2016) or 'negativity towards WBA in the workplace' (Massie & Ali, 2016). There are a number of studies which have highlighted the issues that need attention and innovative solutions. Two comprehensive reviews of the WBA literature (Massie & Ali, 2016; Miller & Archer, 2010) have identified the key problems which have 'negatively impacted on the effectiveness of WBA tools as learning aids' (Massie & Ali, 2016). They are: a limited understanding of the purpose of WBA; too little time to engage in assessment activities; inadequate training of assessors; and limited evidence demonstrating the impact of WBA on learning and patient care. Clarifying the purpose of WBA, addressing the training needs of assessors and the need for research to demonstrate evidence of benefit are discussed elsewhere in this paper. This section specifically focuses on the time constraints relevant to WBA.

There are several studies highlighting the limitations of time and the impact thereof on implementing WBA (Hrisos et al., 2008; Wilkinson et al., 2008; Cohen et al., 2009; Dijksterhuis et al., 2013; Tsagkatakis & Choudhary, 2013; Rees et al., 2014; General Medical Council, 2018). There are also studies describing the difficulties trainees encounter when trying to engage trainers and/or organise WBA activities (McKavanagh et al., 2012; Bindal et al., 2013; Mehta et al., 2013). There are currently two basic approaches to addressing this problem. First, allocating time to conduct training and assessment activities in the work schedules of clinician-educators tasked with the role of supervising trainees. The UK Committee of Postgraduate Dental Deans and Directors (2012) has adopted this approach and stated that allocated trainers 'must have adequate time for training clearly identified in their job plans or appointment systems'. This decision may be making a significant contribution to the more positive attitude towards WBA reported in UK dentistry programmes (Grieverson et al., 2011; Kirton et al., 2013). This approach to the training of doctors has also been proposed by the Academy of Royal Colleges in the UK (2016). While budgetary constraints are likely to be posited as the major reason for limited implementation of the idea (Massie & Ali, 2016), the contribution of train-

ing to improving patient safety, reducing clinical errors and improving patient outcomes should serve as a strong counter argument. Second, health care systems need to be designed in such a way that WBA is an integral part of delivering good patient care rather than a 'stand-alone' assessment activity which competes with the core business of busy clinicians (Hauer et al., 2011). This radical reconceptualization of assessment as both an educational and clinical care issue has already been described (Kogan et al., 2014). Future studies need to focus on strategies which achieve this goal by embedding WBA into employment contracts and the design of health care systems.

13. Digital Technologies and WBA

The rapidly expanding role of digital technologies in education is widely acknowledged and the capacity of such technologies to improve the design, delivery and analysis of education and assessment processes is vast (Care, 2018). Health professions education is a relative late comer to the 21st century digital education arena and much still needs to be learnt about technology-supported information management options for education and assessment in the workplace. Technology can be used to administer, aggregate, and synthesise assessment data to facilitate reflection on performance; provide trainee feedback; inform summative decisions about progression and certification; and compare trainee performance against national or international standards (Lockyer et al., 2017). A recent paper (Lockyer et al., 2017) provides some examples of digital technologies which have already found their way into health professions education. These include Internet-based mini-CEX applications (Ferenchick et al., 2010); mobile technology for administering mini-CEX assessments in the workplace (Ferenchick et al., 2013); digital patient records for clinical training (Hammoud et al., 2012); social networks like wikis for increasing student engagement and feedback (Cheston et al., 2013); virtual patients for assessing the development of clinical reasoning (Cook & Triola, 2009d), electronic learning management systems for tracking and submitting online assessments (Ellaway & Masters, 2008), and intelligence tools for simulation-based and game-based education (Akl et al., 2008, 2010). While these serve as useful examples of small-scale technology-supported learning and assessment practices, more evidence is needed to strengthen the case for large-scale implementation of these methods and strategies (Ruitz et al., 2006; Choules,

2007; Sandars & Haythornthwaite, 2007; Cook et al., 2008; Ellaway & Masters, 2008; Masters & Ellaway, 2008; Cook, 2009a; Frehywot et al., 2013; Masters et al., 2016).

A major criticism of current assessment strategies in medical training programmes is the limited capacity for global up scaling of WBA initiatives by health professions accreditation and licencing bodies. While many small-scale programmes at individual institutions exist, and national postgraduate training programmes, which include WBA, have been rolled out in several countries previously mentioned, these have not been without the significant challenges highlighted in this paper. Although the use of digital technology in WBA programmes can address some of these challenges by improving data entry, online data management, and the use of assessment data to guide further learning, the mandate for large-scale implementation of digital technologies in WBA lies beyond the 'niceties' of such technology. The development of international electronic learning management systems (global consortia) capable of manipulating large sets of educational data (Big Data) are needed (Care, 2018). Internationally, big educational datasets are increasingly being studied using learning analytics and educational data mining to improve teaching and learning processes (Baker & Inventado, 2014; Papamitsiou & Economides, 2014). Learning analytics aims to facilitate the process of data measurement, collection, analysis and reporting for the 'purpose of understanding and optimizing learning' (Cooper, 2012). Educational Data Mining (EDM) focuses on providing insights into teaching practices and student learning using data from assessments and learning experiences (Baker & Yasef, 2009).

Work exploring the use of big educational datasets to influence assessment practices has been slower (Baker & Inventado, 2014) but there are now projects where the use of such data to inform assessment practices are being explored (Wilson & Scalise, 2015). The Assessment and Teaching of 21st century Skills is an example of a large multi-country international project focusing on digital technology-based learning and assessment of a range of skills considered essential for achieving success in the 21st century. These skills include creativity, critical thinking and problem solving, collaborative skills, information technology skills and new forms of literacy, and social, cultural, and metacognitive awareness. (Griffin & Care, 2015). While health professions educators are starting to explore the utility of learning analytics and educational data mining, much work is needed to real-

ize their full potential in mainstream practice (Ellaway et al., 2014; Norcini et al., 2018). In the interim, there is a growing body of education literature in other disciplines describing the actual and predicted contribution of learning analytics and data mining to education programme and assessment development and design (Sin & Muthu, 2015); international educational policy making (Macfadyen et al., 2014; Morrison & van der Werf, 2016); and large-scale international research using educational datasets which are big enough to provide robust answers to, and innovative solutions to the major challenges facing education and assessment (Morrison & van der Werf, 2016). The mandate to exploit such digital technologies to catalyze broader uptake of WBA and support large-scale, data-driven research endeavours in health professions education and assessment is clear.

While such large-scale advances in WBA are still in the future, there are digital technologies which currently make a significant contribution to health professions education. The best example is the use of simulation to provide practical training of health care professionals to enhance their performance in the workplace (Cook et al., 2011; Cook et al., 2013a). In comparison with other methods of instruction, it is associated with small to moderate positive effects (Cook et al., 2012) and there is preliminary evidence suggesting translational benefits such as better patient care and improved health care outcomes (Zendejas, et al., 2013; McGaghie et al., 2014). Unfortunately, validity evidence for simulation-based assessment is sparse and the rigour of existing studies is not optimal (Cook et al., 2013b). Existing studies highlight the limitations of our current knowledge and endorse the statement that ‘further no-intervention-controlled studies or comparisons with traditional instructional methods are not needed’ (Cook, 2009a). Rather, there is need for clarifying how and when to use e-learning through ‘basic science’ research and ‘field tests’ comparing e-learning interventions.

14. Digital Portfolios and WBA

Portfolios, broadly defined as ‘frameworks for collecting, analysing and documenting the successful acquisition of competence and performance’ (Holmboe et al., 2017), have become one of the most popular ways of improving the efficiency and efficacy of WBA data management (Lockyer, et al., 2017). Increasingly digital portfolios are being used

to assess workplace-based learning and to support deliberate practice (Driessen et al., 2007a, 2007b; Buckley et al., 2009; Tochel et al., 2009; van Tartwijk & Driessen, 2009; Moores & Parks, 2010). Universal success with portfolios is, however, not a given and specific issues need attention during implementation, including clearly communicated goals and procedures; integration with curriculum and assessment; flexible structure; support through mentoring; mentor training; and measures to increase feasibility and reduce the time required for completion and assessment (Driessen et al., 2007a, van Tartwijk & Driessen, 2009; Taylor et al., 2014; Heeneman & de Grave, 2017). The ongoing challenges associated with the use of portfolios in postgraduate training in the UK highlight the importance of sufficient time, regular feedback, and user ‘buy-in’ (Taylor et al., 2014; Hrisos et al., 2008).

The use of learning analytics to enhance the educational value of E-portfolios holds promise for the future. Ideally, WBA data collated in an E-portfolio can be used to draw inferences about a trainee’s progress and achievement (Mislevy et al., 2012). A recent study has confirmed the feasibility of using a commercial E-portfolio system to generate personalised feedback for medical trainees using WBA data (Van der Schaaf et al., 2017). In this study the workplace-based tasks were operationalised as units of professional practice (entrustable professional activities) and trainees sent electronic competency requests to supervisors to trigger assessment of their level of entrustment. These data were used to generate personalised feedback based on performance and customised for each EPA. Feedback was provided using both text responses (‘just-in-time’ messages) and visualization of progress using interactive graphics. The authors highlight the importance of end user engagement in the design and development of these systems to ensure ‘buy-in’. This approach holds great promise for improving the key elements of feedback (i.e. performance-based, interactive and immediate) (Ramani & Krakov, 2012; Hattie & Timperley, 2007).

15. The Future of WBA

This paper has provided an update on the current status of WBA in health professions education, including ongoing challenges, and avenues that need further exploration to enhance the quality and uptake of WBA. Key issues that require attention include the following:

- Clarify the purpose of WBA, especially in settings of dual formative and summative use
- Ensure adequate training of users, both supervisors and trainees
- Include users in design and implementation processes to improve utility and engagement
- Increase the use of narrative descriptions to describe performance (qualitative paradigm)
- Use qualitative research criteria to ensure the rigour of narrative data for decision-making
- If used, rating scales should be meaningful to clinician-educators
- Shift the focus of WBA to evaluate the quality of patient care and health outcomes
- Expand the use of digital technology in WBA to reduce user burden
- Embed WBA in employment contracts and the design of health care systems
- Conduct research to explore the meaning, significance and implications of rater variability
- Use learning analytics to improve feedback and enhance the learning experiences of WBA
- Use data mining of Big Data to explore the impact of WBA on patient care

16. Concluding Remarks

Based on this overview of WBA, its challenges and potential ways of moving forward it is clear that there have been extensive developments in this aspect of health professions education over the past 40 years. It is also apparent that we have learnt major lessons along the way. While WBA is highly relevant to the training of competent health care practitioners, assessment in authentic settings remains a challenge. This hurdle needs to be overcome because assessment in the workplace is critical to determining trainees' readiness to be trusted to provide safe, effective and efficient patient care, a public expectation in the 21st century.

17. References

Academy of Medical Royal Colleges (2016). Improving assessment: further guidance and recommendations. London: AoMRC. Available at http://www.aomrc.org.uk/wp-content/uploads/2016/06/Improving_assessment_Further_GR_0616-1.pdf.

- Akl, E. A., Sackett, K. M., Pretorius, R., Bhoopathi, P. S. S., Mustafa, R., Schünemann, H. and Erdley, W. S. (2008). Educational games for health professionals. *Cochrane Database of Systematic Reviews*, (1). <https://doi.org/10.1002/14651858.CD006411.pub2>. PMID: 18254103.
- Akl, E. A., Pretorius, R. W., Sackett, K., Erdley, W. S., Bhoopathi, P. S., Alfarah, Z. and Schünemann, H. J. (2010). The effect of educational games on medical students' learning outcomes: a systematic review: BEME Guide No 14. *Medical Teacher*, 32(1), 16–27. <https://doi.org/10.3109/01421590903473969>. PMID: 20095770.
- Archer, J., Norcini, J., Southgate, L., Heard, S. and Davies, H. (2008). mini-PAT (Peer Assessment Tool): A valid component of a national assessment programme in the UK? *Advances in Health Sciences Education: Theory and Practice*, 13(2), 181–92. <https://doi.org/10.1007/s10459-006-9033-3>. PMID:17036157.
- Baker, R. S. and Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, 1(1), 3–17.
- Baker, R. S. and Inventado, P. S. (2014). Educational Data Mining and Learning Analytics. In: Larusson J., White, B. (Eds.) *Learning Analytics*, Springer, New York, NY. p. 61–75. PMCID: PMC4355053.
- Basu, I., Parvizi, S. and Chin, K. (2013). The perception of online work-based assessments. *The Clinical Teacher*, 10(2), 73–77. <https://doi.org/10.1111/j.1743-498X.2012.00609.x>. PMID: 23480106.
- Batalden, P., Leach, D., Swing, S., Dreyfus, H. and Dreyfus, S. (2002). General competencies and accreditation in graduate medical education. *Health Affairs*, 21(5), 103–11. <https://doi.org/10.1377/hlthaff.21.5.103>. PMID: 12224871.
- Beard, J. and Bussey, M. (2007). Workplace-based assessment. *The Bulletin of the Royal College of Surgeons of England*, 89(5), 158–60. <https://doi.org/10.1308/147363507X196549>.
- Bindal, T., Wall, D. and Goodyear, H. M. (2011). Trainee doctors' views on workplace-based assessments: are they just a tick box exercise? *Medical Teacher*, 33(11), 919–27. <https://doi.org/10.3109/0142159X.2011.558140>. PMID: 22022902.
- Bindal, N., Goodyear, H., Bindal, T. and Wall, D. (2013). DOPS assessment: A study to evaluate the experience and opinions of trainees and assessors. *Medical Teacher*, 35(6), e1230–34. <https://doi.org/10.3109/0142159X.2012.746447>. PMID: 23627359.
- Black, P. and Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: principles, Policy & Practice*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>.
- Bodle, J. F., Kaufmann, S. J., Bisson, D., Nathanson, B. and Binney, D. M. (2008). Value and face validity of objective structured assessment of technical skills (OSATS) for

- work based assessment of surgical skills in obstetrics and gynaecology. *Medical Teacher*, 30(2), 212–16. <https://doi.org/10.1080/01421590701881624>. PMID: 18464149.
- Bok, H. G., Teunissen, P. W., Favier, R. P., Rietbroek, N. J., Theyse, L. F., Brommer, H., Haarhuis, J. C., van Beukelen, P., van der Vleuten, C. P. and Jaarsma, D. A. (2013). Programmatic assessment of competency-based workplace learning: when theory meets practice. *BMC Medical Education*, 13(1), 123. <https://doi.org/10.1186/1472-6920-13-123>. PMID: 24020944, PMCID: PMC3851012.
- Boud, D. (1990). Assessment and the promotion of academic values. *Studies in Higher Education*, 15(1), 101–11. <https://doi.org/10.1080/03075079012331377621>.
- Boulet, J. R., McKinley, D. W., Norcini, J. J. and Whelan, G. P. (2002). Assessing the comparability of standardized patient and physician evaluations of clinical skills. *Advances in Health Sciences Education: Theory and Practice*, 7(2), 85–97. <https://doi.org/10.1023/A:1015750009235>. PMID: 12075142.
- Braddock, C. H., Fihn, S. D., Levinson, W., Jonsen, A. R. and Pearlman, R. A. (1997). How doctors and patients discuss routine clinical decisions: informed decision making in the outpatient setting. *Journal of General Internal Medicine*, 12(6), 339–45. <https://doi.org/10.1007/s11606-006-5081-x>, <https://doi.org/10.1046/j.1525-1497.1997.00057.x>. PMID: 9192250, PMCID: PMC1497128.
- Brinkman, W. B., Geraghty, S. R., Lanphear, B. P., Khoury, J. C., del Rey, J. A. G., DeWitt, T. G. and Britto, M. T. (2007). Effect of multisource feedback on resident communication skills and professionalism: a randomized controlled trial. *Archives of Pediatrics and Adolescent Medicine*, 161(1), 44–49. <https://doi.org/10.1001/archpedi.161.1.44>. PMID: 17199066.
- Brown, S. (2005). Assessment for learning. *Learning and Teaching in Higher Education*, (1), 81–89.
- Buckley, S., Coleman, J., Davison, I., Khan, K. S., Zamora, J., Malick, S., Morley, D., Pollard, D., Ashcroft, T., Popovic, C. and Sayers, J. (2009). The educational effects of portfolios on undergraduate student learning: A Best Evidence Medical Education (BEME) systematic review, BEME Guide No. 11. *Medical Teacher*, 31(4), 282–98. <https://doi.org/10.1080/01421590902889897>. PMID: 19404891.
- Burford, B., Illing, J., Kergon, C., Morrow, G. and Livingston, M. (2010). User perceptions of multi-source feedback tools for junior doctors. *Medical Education*, 44(2), 165–76. <https://doi.org/10.1111/j.1365-2923.2009.03565.x>. PMID: 20059677.
- Care E. (2018). Twenty-First Century Skills: From Theory to Action. In: Care E., Griffin P., Wilson M. (Eds.) *Assessment and Teaching of 21st Century Skills*. Educational Assessment in an Information Age. p. 3–17. Cham: Springer. https://doi.org/10.1007/978-3-319-65368-6_1.
- Carraccio, C., Wolfsthal, S. D., Englander, R., Ferentz, K. and Martin, C. (2002). Shifting paradigms: from Flexner to competencies. *Academic Medicine*, 77(5), 361–67. <https://doi.org/10.1097/00001888-200205000-00003>. PMID: 12010689.
- Carraccio, C. L. and Englander, R. (2013). From Flexner to competencies: reflections on a decade and the journey ahead. *Academic Medicine*, 88(8), 1067–73. <https://doi.org/10.1097/ACM.0b013e318299396f>. PMID: 23807096.
- Castanelli, D. and Kitto, S. (2011). Perceptions, attitudes, and beliefs of staff anaesthetists related to multi-source feedback used for their performance appraisal. *British Journal of Anaesthesia*, 107(3), 372–377. <https://doi.org/10.1093/bja/aer152>. PMID: 21676891.
- Cheston, C. C., Flickinger, T. E. and Chisolm, M. S. (2013). Social media use in medical education: A systematic review. *Academic Medicine*, 88(6), 893–901. <https://doi.org/10.1097/ACM.0b013e31828ff23>. PMID: 23619071.
- Cho, S. P., Parry, D. and Wade, W. (2014). Lessons learnt from a pilot of assessment for learning. *Clinical Medicine*, 14(6), 577–84. <https://doi.org/10.7861/clinmedicine.14-6-577>. PMID: 25468839, PMCID: PMC4954126.
- Choules, A. P. (2007). The use of e-learning in medical education: a review of the current situation. *Postgraduate Medical Journal*, 83(978), 212–16. <https://doi.org/10.1136/pgmj.2006.054189>. PMID: 17403945, PMCID: PMC2600032.
- Cleland, J. A., Knight, L. V., Rees, C. E., Tracey, S. and Bond, C. M. (2008). Is it me or is it them? Factors that influence the passing of underperforming students. *Medical Education*, 42(8), 800–809. <https://doi.org/10.1111/j.1365-2923.2008.03113.x>. PMID: 18715477.
- Cohen, S. N., Farrant, P. B. J. and Taibjee, S. M. (2009). Assessing the assessments: UK dermatology trainees' views of the workplace assessment tools. *British Journal of Dermatology*, 161(1), 34–39. <https://doi.org/10.1111/j.1365-2133.2009.09097.x>. PMID: 19416267.
- Collins, J. (2010). *Foundation for excellence: An evaluation of the Foundation Programme*. London: Medical Education England.
- Cook, D. A., Levinson, A. J., Garside, S., Dupras, D. M., Erwin, P. J. and Montori, V. M. (2008). Internet-based learning in the health professions: a meta-analysis. *Journal of the American Medical Association*, 300(10), 1181–96. <https://doi.org/10.1001/jama.300.10.1181>, <https://doi.org/10.1503/cmaj.071366>. PMCID: PMC2292790.
- Cook, D. A. (2009a). The failure of e-learning research to inform educational practice and what we can do about it. *Medical Teacher*, 31(2), 158–62. <https://doi.org/10.1080/01421590802691393>. PMID: 19330674.

- Cook, D. A., Dupras, D. M., Beckman, T. J., Thomas, K. G. and Pankratz, V. S. (2009b). Effect of rater training on reliability and accuracy of mini-CEX scores: A randomized, controlled trial. *Journal of General Internal Medicine*, 24(1), 74–79. <https://doi.org/10.1007/s11606-008-0842-3>. PMID: 19002533, PMCID: PMC2607488.
- Cook, D. A. and Beckman, T. J. (2009c). Does scale length matter? A comparison of nine-versus five-point rating scales for the mini-CEX. *Advances in Health Sciences Education: Theory and Practice*, 14(5), 655–64. <https://doi.org/10.1007/s10459-008-9147-x>. PMID: 19034679.
- Cook, D. A. and Triola, M. M. (2009d). Virtual patients: a critical literature review and proposed next steps. *Medical Education*, 43(4), 303–11. <https://doi.org/10.1111/j.1365-2923.2008.03286.x>. PMID: 19335571.
- Cook, D. A., Hatala, R., Brydges, R., Zendejas, B., Szostek, J. H., Wang, A. T., Erwin, P. J. and Hamstra, S. J. (2011). Technology-enhanced simulation for health professions education: A systematic review and meta-analysis. *Journal of the American Medical Association*, 306(9), 978–88. <https://doi.org/10.1001/jama.2011.1234>. PMID: 21900138.
- Cook, D. A., Brydges, R., Hamstra, S. J., Zendejas, B., Szostek, J. H., Wang, A. T., Erwin, P. J. and Hatala, R. (2012). Comparative effectiveness of technology-enhanced simulation versus other instructional methods: A systematic review and meta-analysis. *Simulation in Healthcare*, 7(5), 308–20. <https://doi.org/10.1097/SIH.0b013e3182614f95>. PMID: 23032751.
- Cook, D. A., Brydges, R., Zendejas, B., Hamstra, S. J. and Hatala, R. (2013a). Mastery learning for health professionals using technology-enhanced simulation: A systematic review and meta-analysis. *Academic Medicine*, 88(8), 1178–86. <https://doi.org/10.1097/ACM.0b013e31829a365d>. PMID: 23807104.
- Cook, D. A., Brydges, R., Zendejas, B., Hamstra, S. J. and Hatala, R. (2013b). Technology-enhanced simulation to assess health professionals: A systematic review of validity evidence, research methods, and reporting quality. *Academic Medicine*, 88(6), 872–83. <https://doi.org/10.1097/ACM.0b013e31828ffdcf>. PMID: 23619073.
- Cooper, A. (2012). What is analytics? Definition and essential characteristics. *CETIS Analytics Series*, 1(5), 1–10.
- Crossley, J. and Jolly, B. (2012). Making sense of work-based assessment: Ask the right questions, in the right way, about the right things, of the right people. *Medical Education*, 46(1), 28–37. <https://doi.org/10.1111/j.1365-2923.2011.04166.x>. PMID: 22150194.
- Davies, H., Archer, J., Heard, S. and Southgate, L. (2005). Assessment tools for foundation programmes - A practical guide. *British Medical Journal*, 330(7500), s195–96. <https://doi.org/10.1136/bmj.330.7500.s195-a>.
- Davies, H., Archer, J., Southgate, L. and Norcini, J. (2009). Initial evaluation of the first year of the Foundation Assessment Programme. *Medical Education* 43(1), 74–81. <https://doi.org/10.1111/j.1365-2923.2008.03249.x>. PMID: 19141000.
- Dean, B. J. F. and Duggleby, P. M. (2013). Foundation doctors' experience of their training: a questionnaire study. *JRSM Short Reports*, 4(1), 1–7. <https://doi.org/10.1258/shorts.2012.012095>. PMID: 23413407, PMCID: PMC3572662.
- De Lima, A. A., Barrero, C., Baratta, S., Castillo Costa, Y., Bortman, G., Carabajales, J., Conde, D., Galli, A., Degrange, G. and van der Vleuten, C. (2007). Validity, reliability, feasibility and satisfaction of the Mini-Clinical Evaluation Exercise (Mini-CEX) for cardiology residency training. *Medical Teacher*, 29(8), 785–90. <https://doi.org/10.1080/01421590701352261>. PMID: 17917984.
- De Lima, A. A., Conde, D., Costabel, J., Corso, J. and Van der Vleuten, C. (2013). A laboratory study on the reliability estimations of the mini-CEX. *Advances in Health Sciences Education: Theory and Practice*, 18, 5–13. <https://doi.org/10.1007/s10459-011-9343-y>. PMID: 22193944, PMCID: PMC3569586.
- Dennis, A. A., Foy, M. J., Monrouxe, L. V. and Rees, C. E. (2018). Exploring trainer and trainee emotional talk in narratives about workplace-based feedback processes. *Advances in Health Sciences Education: Theory and Practice*, 23(1), 75–93. <https://doi.org/10.1007/s10459-017-9775-0>. PMID: 28456856, PMCID: PMC5801389.
- Dijksterhuis, M. G. K., Schuwirth, L. W. T., Braat, D. D. M., Teunissen, P. W. and Scheele, F. (2013). A qualitative study on trainees' and supervisors' perceptions of assessment for learning in postgraduate medical education. *Medical Teacher*, 35(8), e1396–402. <https://doi.org/10.3109/0142159X.2012.756576>. PMID: 23600668.
- Donato, A. A., Pangaro, L., Smith, C., Rencic, J., Diaz, Y., Mensinger, J. and Holmboe, E. (2008). Evaluation of a novel assessment form for observing medical residents: A randomised, controlled trial. *Medical Education*, 42(12), 1234–42. <https://doi.org/10.1111/j.1365-2923.2008.03230.x>. PMID: 19120955.
- Driessen, E., van der Vleuten, C., Schuwirth, L., van Tartwijk, J. and Vermunt, J. D. H. M. (2005). The use of qualitative research criteria for portfolio assessment as an alternative to reliability evaluation: A case study. *Medical Education*, 39(2), 214–20. <https://doi.org/10.1111/j.1365-2929.2004.02059.x>. PMID: 15679689.
- Driessen, E., Van Tartwijk, J., van der Vleuten, C. and Wass, V. (2007a). Portfolios in medical education: why do they meet with mixed success? A systematic review. *Medical Education*, 41(12), 1224–1233. <https://doi.org/10.1111/j.1365-2923.2007.02944.x>. PMID: 18045373.

- Driessen, E. W., Muijtjens, A. M., van Tartwijk, J. and van der Vleuten, C. P. (2007b). Web-or paper-based portfolios: Is there a difference? *Medical Education*, 41(11), 1067–73. <https://doi.org/10.1111/j.1365-2923.2007.02859.x>. PMID: 17973767.
- Dudek, N. L., Marks, M. B. and Regehr, G. (2005). Failure to fail: the perspective of clinical supervisors. *Academic Medicine*, 80(10 suppl.), S84–87. <https://doi.org/10.1097/00001888-200510001-00023>.
- Durning, S. J., Cation, L. J., Markert, R. J. and Pangaro, L. N. (2002). Assessing the reliability and validity of the mini—clinical evaluation exercise for internal medicine residency training. *Academic Medicine*, 77(9), 900–904. <https://doi.org/10.1097/00001888-200209000-00020>. PMID: 12228088.
- Ellaway, R. and Masters, K. (2008). AMEE Guide 32: e-Learning in medical education Part 1: learning, teaching and assessment. *Medical Teacher*, 30(5), 455–73. <https://doi.org/10.1080/01421590802108331>. PMID: 18576185.
- Ellaway, R. H., Pusic, M. V., Galbraith, R. M. and Cameron, T. (2014). Developing the role of big data and analytics in health professional education. *Medical Teacher*, 36(3), 216–22. <https://doi.org/10.3109/0142159X.2014.874553>. PMID: 24491226.
- Englander, R., Frank, J. R., Carraccio, C., Sherbino, J., Ross, S., Snell, L. and ICBME Collaborators. (2017). Toward a shared language for competency-based medical education. *Medical Teacher*, 39(6), 582–587. <https://doi.org/10.1080/0142159X.2017.1315066>. PMID: 28598739.
- Ferenchick, G. S., Foreback, J., Towfiq, B., Kavanaugh, K., Solomon, D. and Mohmand, A. (2010). The implementation of a mobile problem-specific electronic CEX for assessing directly observed student–patient encounters. *Medical Education online*, 15(1), 4276. <https://doi.org/10.3402/meo.v15i0.4276>. PMID: 20174616, PMCID: PMC2823396.
- Ferenchick, G. S., Solomon, D., Foreback, J., Towfiq, B., Kavanaugh, K., Warbasse, L., Addison, J., Chames, F., Dandan, A. and Mohmand, A. (2013). Mobile technology for the facilitation of direct observation and assessment of student performance. *Teaching and Learning in Medicine*, 25(4), 292–99. <https://doi.org/10.1080/10401334.2013.827972>. PMID: 24112197.
- Finall, A. (2012). Trainers’ perceptions of the direct observation of practical skills assessment in histopathology training: a qualitative pilot study. *Journal of Clinical Pathology*, 65, 538–40. <https://doi.org/10.1136/jclinpath-2012-200682>. PMID: 22447913.
- Fokkema, J. P., Teunissen, P. W., Westerman, M., van der Lee, N., van der Vleuten, C. P., Scherpbier, A. J., Dörr, P. J. and Scheele, F. (2013). Exploration of perceived effects of innovations in postgraduate medical education. *Medical Education*, 47(3), 271–81. <https://doi.org/10.1111/medu.12081>. PMID: 23398013.
- Frank, J. R. and Danoff, D. (2007). The CanMEDS initiative: implementing an outcomes-based framework of physician competencies. *Medical Teacher*, 29(7), 642–47. <https://doi.org/10.1080/01421590701746983>. PMID: 18236250.
- Frank, J. R., Mungroo, R., Ahmad, Y., Wang, M., De Rossi, S. and Horsley, T. (2010a). Toward a definition of competency-based education in medicine: a systematic review of published definitions. *Medical Teacher*, 32(8), 631–37. <https://doi.org/10.3109/0142159X.2010.500898>. PMID: 20662573.
- Frank, J. R., Snell, L. S., Cate, O. T., Holmboe, E. S., Carraccio, C., Swing, S. R., Harris, P., Glasgow, N. J., Campbell, C., Dath, D. & Harden, R. M. (2010b). Competency-based medical education: Theory to practice. *Medical Teacher*, 32(8), 638–45. <https://doi.org/10.3109/0142159X.2010.501190>. PMID: 20662574.
- Frank, J. R., Snell, L., Englander, R., Holmboe, E. S. and ICBME Collaborators. (2017). Implementing competency-based medical education: moving forward. *Medical Teacher*, 568–573. <https://doi.org/10.1080/0142159X.2017.1315069>. PMID: 28598743.
- Frehywot, S., Vovides, Y., Talib, Z., Mikhail, N., Ross, H., Wohltjen, H., Bedada, S., Korhumel, K., Koumare, A. K. and Scott, J. (2013). E-learning in medical education in resource constrained low-and middle-income countries. *Human Resources for Health*, 11(1), 4. <https://doi.org/10.1186/1478-4491-11-4>. PMID: 23379467, PMCID: PMC3584907.
- Frank, J., Chen, L., Bhutta, Z.A., Cohen, J., Crisp, N., Evans, T., Fineberg, H., Garcia, P., Ke, Y., Kelley, P. and Kistnasamy, B. (2010). Health professionals for a new century: transforming education to strengthen health systems in an interdependent world. *The Lancet*, 376 (9756), 1923–58. [https://doi.org/10.1016/S0140-6736\(10\)61854-5](https://doi.org/10.1016/S0140-6736(10)61854-5).
- General Medical Council (2011). Learning and Assessment in The Clinical Environment: The Way Forward. London: GMC.
- General Medical Council (2018). National Training Surveys 2018. London: GMC. Available at https://www.gmc-uk.org/-/media/documents/dc11391-nts-2018-initial-findings-report_pdf-75268532.pdf.
- General Medical Council (2013). Good Medical Practice. Available at: https://www.gmc-uk.org/-/media/documents/good-medical-practice---english-1215_pdf-51527435.pdf.
- Ginsburg, S., Eva, K. and Regehr, G. (2013). Do in-training evaluation reports deserve their bad reputations? A study of the reliability and predictive ability of ITER scores and narrative comments. *Academic Medicine*, 88(10), 1539–44.

- <https://doi.org/10.1097/ACM.0b013e3182a36c3d>. PMID: 23969371.
- Ginsburg, S., van der Vleuten, C., Eva, K. W. and Lingard, L. (2016). Hedging to save face: a linguistic analysis of written comments on in-training evaluation reports. *Advances in Health Sciences Education: Theory and Practice*, 21(1), 175–88. <https://doi.org/10.1007/s10459-015-9622-0>. PMID: 26184115.
- Govaerts, M. J., van der Vleuten, C. P., Schuwirth, L. W. and Muijtens, A. M. (2007). Broadening perspectives on clinical performance assessment: rethinking the nature of in-training assessment. *Advances in Health Sciences Education: Theory and Practice*, 12 (2), 239–260. PMCid: PMC3557509.
- Govaerts, M. & van der Vleuten, C. P. (2013). Validity in work-based assessment: Expanding our horizons. *Medical Education*, 47(12), 1164–74. <https://doi.org/10.1111/medu.12289> PMID: 24206150.
- Govaerts, M. J. B., van de Wiel, M. W. J., Schuwirth, L. W. T., van der Vleuten, C. P. M. & Muijtens, A. M. M. (2013). Workplace-based assessment: Raters' performance theories and constructs. *Advances in Health Sciences Education: Theory and Practice*, 18(3), 375–96.
- Grieverson, B., Kirton, J. A., Palmer, N. O. A. and Balmer, M. C. (2011). Evaluation of workplace based assessment tools in dental foundation training. *British Dental Journal*, 211(4), E8. <https://doi.org/10.1038/sj.bdj.2011.681>. PMID: 21869769.
- Griffin, P. and Care, E. (2015). The ATC21S Method. In: Griffin P., Care E. (Eds.) *Assessment and Teaching of 21st Century Skills*. Educational Assessment in an Information Age, Dordrecht: Springer. p. 3–33. https://doi.org/10.1007/978-94-017-9395-7_1.
- Guba, E. G. and Lincoln, Y. S. (1989). What is this constructivist paradigm anyway? In Guba, E. A. & Lincoln Y. S. (Eds.). *Fourth Generation Evaluation*, Newbury Park, California: Sage. p. 79–117. PMCid: PMC304045.
- Hammoud, M. M., Dalrymple, J. L., Christner, J. G., Stewart, R. A., Fisher, J., Margo, K., Ali, I. I., Briscoe, G. W. and Pangaro, L.N. (2012). Medical student documentation in electronic health records: a collaborative statement from the Alliance for Clinical Education. *Teaching and Learning in Medicine*, 24(3), 257–66. <https://doi.org/10.1080/10401334.2012.692284>. PMID: 22775791.
- Harden, R. M. (1999). AMEE Guide No. 14: Outcome-based education: Part 1- An introduction to outcome-based education. *Medical Teacher*, 21(1), 7–14.
- Harris, P., Bhanji, F., Topps, M., Ross, S., Lieberman, S., Frank, J. R., Snell, L., Sherbino, J. & ICBME Collaborators. (2017). Evolving concepts of assessment in a competency-based world. *Medical Teacher*, 39(6), p. 603–08. <https://doi.org/10.1080/0142159X.2017.1315071>. PMID: 28598736.
- Hatala, R. and Norman, G. R. (1999). In-training evaluation during an internal medicine clerkship. *Academic Medicine*, 74(10), S118–20. <https://doi.org/10.1097/00001888-199910000-00059>.
- Hattie, J. and Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>.
- Hauer, K. E. (2000). Enhancing feedback to students using the mini-CEX (Clinical Evaluation Exercise). *Academic Medicine*, 75(5), 524. <https://doi.org/10.1097/00001888-200005000-00046>. PMID: 10824798.
- Hauer, K. E., Holmboe, E. S. and Kogan, J. R. (2011). Twelve tips for implementing tools for direct observation of medical trainees' clinical skills during patient encounters. *Medical Teacher*, 33(1), 27–33. <https://doi.org/10.3109/0142159X.2010.507710>. PMID: 20874011.
- Heeneman, S. and de Grave, W. (2017). Tensions in mentoring medical students toward self-directed and reflective learning in a longitudinal portfolio-based mentoring system - An activity theory analysis. *Medical Teacher*, 39(4), 368–76. <https://doi.org/10.1080/0142159X.2017.1286308>. PMID: 28379081.
- Hodges, B. (2013). Assessment in the post-psychometric era: Learning to love the subjective and collective. *Medical Teacher*, 35(7), 564–68. <https://doi.org/10.3109/0142159X.2013.789134>. PMID: 23631408.
- Holmboe, E. S., Hawkins, R. E. and Huot, S. J. (2004). Effects of training in direct observation of medical residents' clinical competence. *Annals of Internal Medicine*, 140(11), 874–81. <https://doi.org/10.7326/0003-4819-140-11-200406010-00008>. PMID: 15172901.
- Holmboe, E.E., Davis, H.M and Carraccio, C. (2017). Portfolios. In Holmboe, E. S., Hawkins, R. E. (Eds.). *Practical Guide to the Evaluation of Clinical Competence*. 2nd ed., Philadelphia: Elsevier Health Sciences. p. 270–87.
- Holmboe, E. S., Sherbino, J., Long, D. M., Swing, S. R., Frank, J. R. and International CBME Collaborators. (2010). The role of assessment in competency-based medical education. *Medical Teacher*, 32(8), 676–82. <https://doi.org/10.3109/0142159X.2010.500704>. PMID: 20662580.
- Holmboe, E. S., Ward, D. S., Reznick, R. K., Katsufakis, P. J., Leslie, K. M., Patel, V. L., Ray, D. D. and Nelson, E. A. (2011). Faculty development in assessment: The missing link in competency-based medical education. *Academic Medicine*, 86(4), 460–67. <https://doi.org/10.1097/ACM.0b013e31820cb2a7>. PMID: 21346509.
- Holmboe, E.S. and Harden, R.M. (2017). Outcome-based education. In: Dent, J.mA., Harden, R. M., Hunt, D. (Eds.). A

- Practical Guide for Medical Teachers, 5th ed., Edinburgh: Elsevier. p. 114–21.
- Hrisos, S., Illing, J. C. and Burford, B. C. (2008). Portfolio learning for foundation doctors: Early feedback on its use in the clinical workplace. *Medical Education*, 42(2), 214–23. <https://doi.org/10.1111/j.1365-2923.2007.02960.x>. PMID:18230095.
- Isles, M. and Pugsley, L. (2010). Surgical procedure-based assessment: how did we get here and where are we going? *The Bulletin of the Royal College of Surgeons of England*, 92(7), 1–3. <https://doi.org/10.1308/147363510X511625>.
- Jenkins, L., Mash, B. and Derese, A. (2013). The national portfolio of learning for postgraduate family medicine training in South Africa: experiences of registrars and supervisors in clinical practice. *BMC Medical Education*, 13(1), 149. <https://doi.org/10.1186/1472-6920-13-149>. PMID:24207009, PMCID: PMC4226197.
- Jones, R., Higgs, R., De Angelis, C. and Prideaux, D. (2001). Changing face of medical curricula. *The Lancet*, 357(9257), 699–703. [https://doi.org/10.1016/S0140-6736\(00\)04134-9](https://doi.org/10.1016/S0140-6736(00)04134-9).
- Jurd, S., de Beer, W., Aimer, M., Fletcher, S., Halley, E., Schapper, C. and Orkin, M. (2015). Introducing a competency based fellowship programme for psychiatry in Australia and New Zealand. *Australasian Psychiatry*, 23(6), 699–705. <https://doi.org/10.1177/1039856215600898>. PMID: 26400448.
- Jyothirmayi, R. (2012). Case-based discussion: assessment tool or teaching aid? *Clinical Oncology*, 24(10), 649–53. <https://doi.org/10.1016/j.clon.2012.07.008>. PMID:22892380.
- Kirton, J., Palmer, N., Grieveson, B. and Balmer, M. (2013). A national evaluation of workplace-based assessment tools (WPBAs) in foundation dental training: A UK study. Effective and useful but do they provide an equitable training experience? *British Dental Journal*, 214(6), 305–09. <https://doi.org/10.1038/sj.bdj.2013.302>. PMID:23518981.
- Kogan, J. R., Holmboe, E. S. and Hauer, K. E. (2009). Tools for direct observation and assessment of clinical skills of medical trainees: a systematic review. *Journal of the American Medical Association*, 302(12), 1316–26. <https://doi.org/10.1001/jama.2009.1365>. PMID:19773567.
- Kogan, J. R., Hess, B. J., Conforti, L. N. and Holmboe, E. S. (2010). What drives faculty ratings of residents' clinical skills? The impact of faculty's own clinical skills. *Academic Medicine*, 85(Suppl. 10), S25–28. <https://doi.org/10.1097/ACM.0b013e3181ed1aa3>. PMID: 20881697.
- Kogan, J. R., Conforti, L., Bernabeo, E., Iobst, W. and Holmboe, E. (2011). Opening the black box of clinical skills assessment via observation: A conceptual model. *Medical Education*, 45(10), 1048–60. <https://doi.org/10.1111/j.1365-2923.2011.04025.x>. PMID:21916943.
- Kogan, J.R. and Holmboe, E. (2013). Realizing the promise and importance of performance-based assessment. *Teaching and Learning in Medicine*, 25(Suppl. 1), S68–S74. <https://doi.org/10.1080/10401334.2013.842912>. PMID:24246110.
- Kogan, J. R., Conforti, L. N., Iobst, W. F. and Holmboe, E. S. (2014). Reconceptualizing variable rater assessments as both an educational and clinical care problem. *Academic Medicine*, 89(5), 721–27. <https://doi.org/10.1097/ACM.0000000000000221>. PMID: 24667513.
- Lincoln, Y. S. and Guba, E. G. (1985). Establishing trustworthiness. In: Lincoln, Y. S., Guba, E. A. (Eds.). *Naturalistic Inquiry*. Beverly Hills, California: Sage. p. 289–332.
- Lockyer, J., Violato, C. and Fidler, H. (2003). Likelihood of change: a study assessing surgeon use of multisource feedback data. *Teaching and Learning in Medicine*, 15(3), 168–74. https://doi.org/10.1207/S15328015TLM1503_04. PMID:12855387
- Lockyer, J., Carraccio, C., Chan, M.K., Hart, D., Smee, S., Touchie, C., Holmboe, E.S., Frank, J.R. and ICBME Collaborators. (2017). Core principles of assessment in competency-based medical education. *Medical Teacher*, 39(6), 609–16. <https://doi.org/10.1080/0142159X.2017.1315082>. PMID: 28598746.
- Lörwald, A. C., Lahner, F. M., Greif, R., Berendonk, C., Norcini, J. and Huwendiek, S. (2018). Factors influencing the educational impact of Mini-CEX and DOPS: a qualitative synthesis. *Medical Teacher*, 40(4), 414–20. <https://doi.org/10.1080/0142159X.2017.1408901>. PMID: 29188739.
- Lurie, S. J., Mooney, C. J. and Lyness, J. M. (2009). Measurement of the general competencies of the Accreditation Council for Graduate Medical Education: a systematic review. *Academic Medicine*, 84(3), 301–09. <https://doi.org/10.1097/ACM.0b013e3181971f08>. PMID:19240434.
- Maatsch, J.L., Huang, R., Downing, S. and Barker, B. (1983). Predictive validity of medical specialist examinations. Final report for Grant HS 02038-04, National Center of Health Services Research. Office of Medical Education Research and Development, Michigan State University, East Lansing, MI. PMID:6631930.
- Macfadyen, L. P., Dawson, S., Pardo, A. and Gašević, D. (2014). Embracing big data in complex educational systems: The learning analytics imperative and the policy challenge. *Research and Practice in Assessment*, 9, 17–28.
- Makary, M. A. and Daniel, M. (2016). Medical error—the third leading cause of death in the US. *British Medical Journal (Online)*, 353, i2139. <https://doi.org/10.1136/bmj.i2139>.
- Malhotra, S., Hatala, R. and Courneya, C. A. (2008). Internal medicine residents' perceptions of the mini-clinical evaluation exercise. *Medical Teacher*, 30(4), 414–19. <https://doi.org/10.1080/01421590801946962>. PMID:18569664.

- Margolis, M. J., Clauser, B. E., Cuddy, M. M., Ciccone, A., Mee, J., Harik, P. and Hawkins, R. E.. (2006). Use of the mini clinical evaluation exercise to rate examinee performance on a multiple-station clinical skills examination: A validity study. *Academic Medicine*, 81(10 Suppl.), S56–60. <https://doi.org/10.1097/01.ACM.0000236514.53194.f4>. PMID:17001137.
- Marriott, J., Purdie, H., Crossley, J. and Beard, J. D. (2011). Evaluation of procedure-based assessment for assessing trainees' skills in the operating theatre. *British Journal of Surgery*, 98(3), 450–57. <https://doi.org/10.1002/bjs.7342>. PMID:21254025.
- Marshall, G.N. and Hays, R.D. (1994). The Patient Satisfaction Questionnaire Short Form (PSQ-18). Santa Monica, CA: RAND Corporation.
- Martin, J. A., Regehr, G., Reznick, R., Macrae, H., Murnaghan, J., Hutchison, C. and Brown, M. (1997). Objective Structured Assessment Of Technical Skill (OSATS) for surgical residents. *British Journal of Surgery*, 84(2), 273–78. <https://doi.org/10.1046/j.1365-2168.1997.02502.x>, <https://doi.org/10.1002/bjs.1800840237>. PMID:9052454.
- Massie, J. and Ali, J. M. (2016). Workplace-based assessment: a review of user perceptions and strategies to address the identified shortcomings. *Advances in Health Sciences Education: Theory and Practice*, 21(2), 455–73. <https://doi.org/10.1007/s10459-015-9614-0>. PMID: 26003590.
- Masters, K. and Ellaway, R. (2008). e-Learning in medical education. AMEE Guide 32 Part 2: technology, management and design. *Medical Teacher*, 30(5), 474–89. <https://doi.org/10.1080/01421590802108349>. PMID: 18576186.
- Masters, K., Ellaway, R. H., Topps, D., Archibald, D. and Hogue, R. J. (2016). Mobile technologies in medical education: AMEE Guide No. 105. *Medical Teacher*, 38(6), 537–49. <https://doi.org/10.3109/0142159X.2016.1141190>. PMID: 27010681.
- McGaghie, W. C., Issenberg, S. B., Barsuk, J. H. and Wayne, D. B. (2014). A critical review of simulation-based mastery learning with translational outcomes. *Medical Education*, 48(4), 375–85. <https://doi.org/10.1111/medu.12391>. PMID: 24606621.
- McGill, D. A., van der Vleuten, C. P. M. and Clarke, M. J. (2013). A critical evaluation of the validity and the reliability of global competency constructs for supervisor assessment of junior medical trainees. *Advances in Health Sciences Education: Theory and Practice*, 18(4), 701–25. <https://doi.org/10.1007/s10459-012-9410-z>. PMID: 23053869.
- McKavanagh, P., Smyth, A. and Carragher, A. (2012). Hospital consultants and workplace based assessments: how foundation doctors view these educational interactions? *Postgraduate Medical Journal*, 88(1037), 119–24. <https://doi.org/10.1136/postgradmedj-2011-130121>. PMID: 22247316.
- McKee, R. F. (2008). The Intercollegiate Surgical Curriculum Programme (ISCP). *Surgery (Oxford)*, 26(10), 411–16. <https://doi.org/10.1016/j.mpsur.2008.08.007>.
- Mehta, F., Brown, J. and Shaw, N. J. (2013). Do trainees value feedback in case-based discussion assessments? *Medical Teacher*, 35(5), e1166–72. <https://doi.org/10.3109/0142159X.2012.731100>. PMID: 23137264.
- Miller, A. and Archer, J. (2010). Impact of workplace-based assessment on doctors' education and performance: A systematic review. *British Medical Journal*, 341, c5064. <https://doi.org/10.1136/bmj.c5064>. PMID: 20870696, PMCID: PMC2945627.
- Miller, G. E. (1990). The assessment of clinical skills/competence/performance. *Academic Medicine*, 65(Suppl. 9), S63–67. <https://doi.org/10.1097/00001888-199009000-00045>.
- Mislevy, R. J., Behrens, J. T., Dicerbo, K. E. and Levy, R. (2012). Design and discovery in educational assessment: Evidence-centered design, psychometrics, and educational data mining. *Journal of Educational Data Mining*, 4(1), 11–48.
- Mitka M. (2008). Many physician practices fall short on accurate blood pressure measurement. *Journal of the American Medical Association*, 299(24), 2842–44. <https://doi.org/10.1001/jama.299.24.2842>. PMID: 18577720.
- Moonen-van Loon, J. M. W., Overeem, K., Donkers, H. H. L. M., Van der Vleuten, C. P. M. and Driessen, E. W. (2013). Composite reliability of a workplace-based assessment toolbox for postgraduate medical education. *Advances in Health Sciences Education: Theory and Practice*, 18(5), 1087–102. <https://doi.org/10.1007/s10459-013-9450-z>. PMID: 23494202.
- Moores, A. and Parks, M. (2010). Twelve tips for introducing E-Portfolios with undergraduate students. *Medical Teacher*, 32(1), 46–49. <https://doi.org/10.3109/01421590903434151>. PMID: 20095774.
- Morris, A., Hewitt, J. and Roberts, C. M. (2006). Practical experience of using directly observed procedures, mini clinical evaluation examinations, and peer observation in pre-registration house officer (FY1) trainees. *Postgraduate Medical Journal*, 82(966), 285–88. <https://doi.org/10.1136/pgmj.2005.040477>. PMID: 16597818, PMCID: PMC2579636.
- Morrison, K. and van der Werf, G. (2016). Large-scale data, “wicked problems”, and “what works” for educational policy making. *Educational Research and Evaluation*, 22(5-6), 255–59. <https://doi.org/10.1080/13803611.2016.1259789>.

- Morse, J. M. (2015). Critical analysis of strategies for determining rigor in qualitative inquiry. *Qualitative Health Research*, 25(9), 1212–22. <https://doi.org/10.1177/1049732315588501>. PMID: 26184336.
- Murphy, D. J., Bruce, D. A., Mercer, S. W. and Eva, K. W. (2009). The reliability of workplace-based assessment in postgraduate medical education and training: a national evaluation in general practice in the United Kingdom. *Advances in Health Sciences Education: Theory and Practice*, 14(2), 219–32. <https://doi.org/10.1007/s10459-008-9104-8>. PMID: 18306052.
- Nair, B. R., Alexander, H. G., McGrath, B. P., Parvathy, M. S., Kilsby, E. C., Wenzel, J., Frank, I. B., Pachev, G. S. and Page, G. G. (2008). The mini clinical evaluation exercise (mini-CEX) for assessing clinical performance of international medical graduates. *Medical Journal of Australia*, 189(3), 159–61. PMID: 18673104.
- Norcini, J. J., Blank, L. L., Arnold, G. K. and Kimball, H. R. (1995). The mini-CEX (clinical evaluation exercise): a preliminary investigation. *Annals of Internal Medicine*, 123(10), 795–99. <https://doi.org/10.7326/0003-4819-123-10-199511150-00008>. PMID: 7574198.
- Norcini J.J., Blank L.L., Arnold G.K., Kimball H.R. (1997). Examiner differences in the mini-CEX. In: Scherpbier, A. J. J. A., van der Vleuten, C. P. M., Rethans, J. J., van der Steeg, A. F. W. (Eds.), *Advances in Medical Education*. Dordrecht: Springer. p. 170–72. https://doi.org/10.1007/978-94-011-4886-3_49.
- Norcini, J. J., Blank, L. L., Duffy, F. D. and Fortna, G. S. (2003). The mini-CEX: a method for assessing clinical skills. *Annals of Internal Medicine*, 138(6), 476–81. <https://doi.org/10.7326/0003-4819-138-6-200303180-00012>. PMID: 12639081.
- Norcini, J. J. (2005). Current perspectives in assessment: the assessment of performance at work. *Medical Education*, 39(9), 880–89. <https://doi.org/10.1111/j.1365-2929.2005.02182.x>. PMID: 16150027.
- Norcini, J. and Burch, V. (2007). Workplace-based assessment as an educational tool: AMEE Guide No. 31. *Medical Teacher*, 29, 855–71. <https://doi.org/10.1080/01421590701775453>. PMID: 18158655.
- Norcini, J. (2014). Workplace assessment. In Swanick, T. (Ed.). *Understanding Medical Education: Evidence, Theory and Practice* (2nd ed.). London: John Wiley & Sons, Ltd. p. 279–291.
- Norcini, J., Anderson, M. B., Bollela, V., Burch, V., Costa, M. J., Duvivier, R., Hays, R., Palacios Mackay, M. F., Roberts, T. and Swanson, D. (2018). Consensus framework for good assessment. *Medical Teacher*. Published Ahead of Print. Available at <https://www.tandfonline.com/doi/full/10.1080/0142159X.2018.1500016>, <https://doi.org/10.1080/0142159X.2018.1500016>.
- O'Brien, B. C., Harris, I. B., Beckman, T. J., Reed, D. A. and Cook, D. A. (2014). Standards for reporting qualitative research: a synthesis of recommendations. *Academic Medicine*, 89(9), 1245–51. <https://doi.org/10.1097/ACM.0000000000000388>. PMID: 24979285.
- O'Leary, D., Al-Taiar, H., Brown, N., Bajorek, T., Ghazirad, M. and Shaddel, F. (2016). Workplace assessment in crisis? The way forward. *British Journal of Psychiatry Bulletin*, 40(2), 61–63. <https://doi.org/10.1192/pb.bp.114.050153>.
- Paauw, D. S., Wenrich, M. D., Curtis, J. R., Carline, J. D. and Ramsey, P. G. (1995). Ability of primary care physicians to recognize physical findings associated with HIV infection. *Journal of the American Medical Association*, 274(17), 1380–82. <https://doi.org/10.1001/jama.1995.03530170060033>, <https://doi.org/10.1001/jama.274.17.1380>. PMID: 7563564.
- Papamitsiou, Z. and Economides, A. A. (2014). Learning analytics and educational data mining in practice: a systematic literature review of empirical evidence. *Journal of Educational Technology and Society*, 17(4), 49–64.
- Parry-Smith, W., Mahmud, A., Landau, A. and Hayes, K. (2014). Workplace-based assessment: a new approach to existing tools. *The Obstetrician and Gynaecologist*, 16(4), 281–85. <https://doi.org/10.1111/tog.12133>.
- Pelgrim, E. A., Kramer, A. W. M., Mokkink, H. G., Van den Elsen, L., Grol, R. P. T. M. and Van der Vleuten, C. P. M. (2011). In-training assessment using direct observation of single-patient encounters: a literature review. *Advances in Health Sciences Education: Theory and Practice*, 16(1), 131–42. <https://doi.org/10.1007/s10459-010-9235-6>. PMID: 20559868, PMCID: PMC3074070.
- Pelgrim, E. A., Kramer, A. W., Mokkink, H. G. and van der Vleuten, C. P. (2012). The process of feedback in workplace-based assessment: Organisation, delivery, continuity. *Medical Education*, 46(6), 604–12. <https://doi.org/10.1111/j.1365-2923.2012.04266.x>. PMID: 22626052.
- Pereira, E. A. and Dean, B. J. (2009). British surgeons' experiences of mandatory online workplace-based assessment. *Journal of the Royal Society of Medicine*, 102(7), 287–93. <https://doi.org/10.1258/jrsm.2009.080398>. PMID: 19605860, PMCID: PMC2711198.
- Ponnamperuma, G. (2013). Workplace based assessment. In: Walsh, E. (Ed.). *Oxford Textbook of Medical Education*. 1st ed. Oxford: Oxford University Press. p. 537–48. <https://doi.org/10.1093/med/9780199652679.003.0046>.
- Powell, H., DiMarco, A. and Saeed, S. (2014). Trainee and trainer opinion on increasing numbers of workplace-based assessments. *Bulletin of the Royal College of Surgeons*

- of England, 96(5), 160–62. <https://doi.org/10.1308/rcs-bull.2014.96.5.160>.
- Ramani, S. and Krackov, S. K. (2012). Twelve tips for giving feedback effectively in the clinical environment. *Medical Teacher*, 34(10), 787–91. <https://doi.org/10.3109/0142159X.2012.684916>. PMID: 22730899.
- Ramsey, P. G., Wenrich, M. D., Carline, J. D., Inui, T. S., Larson, E. B. and LoGerfo, J. P. (1993). Use of peer ratings to evaluate physician performance. *Journal of the American Medical Association*, 269(13), 1655–60. <https://doi.org/10.1001/jama.269.13.1655>, <https://doi.org/10.1001/jama.1993.03500130069034>. PMID: 8240483.
- Rees, C. E., Knight, L. V. and Cleland, J. A. (2009). Medical educators' metaphoric talk about their assessment relationships with students: 'you don't want to sort of be the one who sticks the knife in them'. *Assessment and Evaluation in Higher Education*, 34(4), 455–67. <https://doi.org/10.1080/02602930802071098>.
- Rees, C. E., Cleland, J. A., Dennis, A., Kelly, N., Mattick, K. and Monrouxe, L. V. (2014). Supervised learning events in the Foundation Programme: A UK-wide narrative interview study. *BMJ Open*, 4(10), e005980. <https://doi.org/10.1136/bmjopen-2014-005980> PMID: 25324323, PMCID: PMC4202004.
- Rethans, J. J., Norcini, J. J., Baron-Maldonado, M., Blackmore, D., Jolly, B. C., LaDuca, T., Lew, S., Page, G. G. and Southgate, L. H. (2002). The relationship between competence and performance: implications for assessing practice performance. *Medical Education*, 36(10), 901–09. <https://doi.org/10.1046/j.1365-2923.2002.01316.x>. PMID:12390456.
- Ringsted, C., Østergaard, D. and Scherpbier, A. J. J. A. (2003). Embracing the new paradigm of assessment in residency training: an assessment programme for first-year residency training in anaesthesiology. *Medical Teacher*, 25(1), 54–62. <https://doi.org/10.1080/0142159021000061431>. PMID: 14741860.
- Roberts, C., Newble, D. I. and O'Rourke, A. (2002). Portfolio-based assessments in medical education: are they valid and reliable for summative purposes? *Medical Education*, 36(10), 899–900. <https://doi.org/10.1046/j.1365-2923.2002.01288.x>. PMID: 12390455.
- Ruitz, J. G., Mintzer, M. J. and Leipzig, R. M. (2006). The impact of e-learning in medical education. *Academic Medicine*, 81(3), 207–12. <https://doi.org/10.1097/00001888-200603000-00002>. PMID: 16501260.
- Ryland, I., Brown, J., O'Brien, M., Graham, D., Gillies, R., Chapman, T. and Shaw, N. (2006). The portfolio: how was it for you? Views of F2 doctors from the Mersey Deanery Foundation Pilot. *Clinical Medicine*, 6(4), 378–80. <https://doi.org/10.7861/clinmedicine.6-4-378>. PMID: 16956144, PMCID: PMC4953521.
- Sabey, A. and Harris, M. (2011). Training in hospitals: what do GP specialist trainees think of workplace-based assessments? *Education for Primary Care*, 22(2), 90–99. <https://doi.org/10.1080/14739879.2011.11493974>. PMID: 21439140.
- Saedon, H., Salleh, S., Balakrishnan, A., Imray, C. H. and Saedon, M. (2012). The role of feedback in improving the effectiveness of workplace based assessments: A systematic review. *BMC Medical Education*, 12(1), 25. <https://doi.org/10.1186/1472-6920-12-25>. PMID: 22551353, PMCID: PMC3432628.
- Sanders, J. and Haythornthwaite, C. (2007). New horizons for e-learning in medical education: ecological and Web 2.0 perspectives. *Medical Teacher*, 29(4), 307–10. <https://doi.org/10.1080/01421590601176406>. PMID: 17786742.
- Sargeant, J. M., Mann, K. V., Ferrier, S. N., Langille, D. B., Muirhead, P. D., Hayes, V. M. and Sinclair, D. E. (2003). Responses of rural family physicians and their colleague and coworker raters to a multi-source feedback process: a pilot study. *Academic Medicine*, 78(10), S42–44. <https://doi.org/10.1097/00001888-200310001-00014>.
- Sargeant, J., Mann, K. and Ferrier, S. (2005). Exploring family physicians' reactions to multisource feedback: perceptions of credibility and usefulness. *Medical Education*, 39(5), 497–504. <https://doi.org/10.1111/j.1365-2929.2005.02124.x>. PMID: 15842684.
- Scheele, F., Teunissen, P., Luijk, S.V., Heineman, E., Fluit, L., Mulder, H., Meininger, A., Wijnen-Meijer, M., Glas, G., Sluiter, H. & Hummel, T. (2008). Introducing competency-based postgraduate medical education in the Netherlands. *Medical Teacher*, 30(3), 248–53. <https://doi.org/10.1080/01421590801993022>. PMID: 18484450.
- Schuwirth, L. W. and van der Vleuten, C. P. (2006). A plea for new psychometric models in educational assessment. *Medical Education*, 40(4), 296–300. <https://doi.org/10.1111/j.1365-2929.2006.02405.x>. PMID: 16573664.
- Schuwirth, L. W. and van der Vleuten, C. P. (2011). Programmatic assessment: from assessment of learning to assessment for learning. *Medical Teacher*, 33(6), 478–85. <https://doi.org/10.3109/0142159X.2011.565828>. PMID: 21609177.
- Schuwirth, L. and Ash, J. (2013). Assessing tomorrow's learners: in competency-based education only a radically different holistic method of assessment will work. Six things we could forget. *Medical Teacher*, 35(7), 555–59. <https://doi.org/10.3109/0142159X.2013.787140>. PMID: 23641916.
- Sebok-Syer, S. S., Klinger, D. A., Sherbino, J. and Chan, T. M. (2017). Mixed messages or miscommunication? Investigating the relationship between assessors' workplace-based assessment scores and written comments. *Academic Medicine*, 92(12), 1774–79. <https://doi.org/10.1097/ACM.0000000000001743>. PMID: 28562452.

- Shute, V.J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153–89. <https://doi.org/10.3102/0034654307313795>.
- Simmons, M. (2013). How are workplace-based assessments viewed by UK psychiatry trainees? *Psychiatra Danubina*, 25(Suppl. 2), S182–84. PMID: 23995172.
- Sin, K. and Muthu, L. (2015). Application of big data in education data mining and learning analytics – A literature review. *ICTACT Journal on Soft Computing*, 5(4). <https://doi.org/10.21917/ijsc.2015.0145>.
- Swanwick, T. and Chana, N. (2009). Workplace-based assessment. *British Journal of Hospital Medicine*, 70(5), 290–93. <https://doi.org/10.12968/hmed.2009.70.5.42235>. PMID: 19451876.
- Swing, S. R. (2007). The ACGME outcome project: retrospective and prospective. *Medical Teacher*, 29(7), 648–54. <https://doi.org/10.1080/01421590701392903>. PMID: 18236251.
- Taylor, A., Dubrey, S. and Das, S. (2014). Opinions of the ePortfolio and workplace-based assessments: A survey of core medical trainees and their supervisors. *Clinical Medicine*, 14(4), 510–16. <https://doi.org/10.7861/clinmedicine.14-5-510>. PMID: 25301912, PMCID: PMC4951960.
- Ten Cate, O. (2005). Entrustability of professional activities and competency-based training. *Medical Education*, 39(12), 1176–77. <https://doi.org/10.1111/j.1365-2929.2005.02341.x>. PMID: 16313574.
- Tochel, C., Haig, A., Hesketh, A., Cadzow, A., Beggs, K., Colthart, I. and Peacock, H. (2009). The effectiveness of portfolios for post-graduate assessment and education: BEME Guide No 12. *Medical Teacher*, 31(4), 299–318. <https://doi.org/10.1080/01421590902883056>. PMID: 19404890.
- Tsagkatakis, M. and Choudhary, A. (2013). Mersey deanery ophthalmology trainees' views of the objective assessment of surgical and technical skills (OSATS) workplace-based assessment tool. *Perspectives on Medical Education*, 2(1), 21–27. <https://doi.org/10.1007/s40037-013-0041-8>. PMID: 23670654, PMCID: PMC3576483.
- Turnbull, J., MacFadyen, J., Van Barneveld, C. and Norman, G. (2000). Clinical work sampling. *Journal of General Internal Medicine*, 15(8), 556–61. <https://doi.org/10.1046/j.1525-1497.2000.06099.x>. PMID: 10940147, PMCID: PMC1495580.
- UK Committee of Postgraduate Dental Deans and Directors. (2012). Quality development of postgraduate dental training in the UK 2012. UK Committee of Postgraduate Dental Deans and Directors. Available at <http://www.copdend.org/guidance/quality-development-of-postgraduate-dental-training-in-the-uk-2012/>.
- Van der Vleuten, C. P. (1996). The assessment of professional competence: developments, research and practical implications. *Advances in Health Sciences Education: Theory and Practice*, 1(1), 41–67. <https://doi.org/10.1007/BF00596229>. PMID: 24178994.
- Van der Vleuten, C. P. and Schuwirth, L. W. (2005). Assessing professional competence: from methods to programmes. *Medical Education*, 39(3), 309–17. <https://doi.org/10.1111/j.1365-2929.2005.02094.x>. PMID: 15733167.
- Van der Vleuten, C. P., Schuwirth, L. W. T., Driessen, E. W., Dijkstra, J., Tigelaar, D., Baartman, L. K. J. and van Tartwijk, J. (2012). A model for programmatic assessment fit for purpose. *Medical Teacher*, 34(3), 205–14. <https://doi.org/10.3109/0142159X.2012.652239>. PMID: 22364452.
- Van der Vleuten, C. and Verhoeven, B. (2013). In-training assessment developments in postgraduate education in Europe. *ANZ Journal of Surgery*, 83(6), 454–59. <https://doi.org/10.1111/ans.12190>. PMID: 23627525.
- Van der Schaaf, M., Donkers, J., Slof, B., Moonen-van Loon, J., van Tartwijk, J., Driessen, E., Badii, A., Serban, O. and Ten Cate, O. (2017). Improving workplace-based assessment and feedback by an E-portfolio enhanced with learning analytics. *Educational Technology Research and Development*, 65(2), 359–80. <https://doi.org/10.1007/s11423-016-9496-8>.
- Van Tartwijk, J. and Driessen, E. W. (2009). Portfolios for assessment and learning: AMEE Guide no. 45. *Medical Teacher*, 31(9), 790–801. <https://doi.org/10.1080/01421590903139201>. PMID: 19811183.
- Veloski, J., Boex, J. R., Grasberger, M. J., Evans, A. and Wolfson, D. B. (2006). Systematic review of the literature on assessment, feedback and physicians' clinical performance: BEME Guide No. 7. *Medical Teacher*, 28(2), 117–28. <https://doi.org/10.1080/01421590600622665>. PMID: 16707292.
- Violato, C., Lockyer, J. M. and Fidler, H. (2008). Changes in performance: A 5-year longitudinal study of participants in a multi-source feedback programme. *Medical Education*, 42(10), 1007–13. <https://doi.org/10.1111/j.1365-2923.2008.03127.x>. PMID: 18823520.
- Vukanovic-Criley, J. M., Criley, S., Warde, C. M., Boker, J. R., Guevara-Matheus, L., Churchill, W. H., Nelson, W. P. and Criley, J. M. (2006). Competency in cardiac examination skills in medical students, trainees, physicians, and faculty: A multicenter study. *Archives of Internal Medicine*, 166(6), 610–16. <https://doi.org/10.1001/archinte.166.6.610>. PMID: 16567598.
- Watling, C., Kenyon, C. F., Schulz, V., Goldszmidt, M. A., Zibrowski, E. and Lingard, L. (2010). An exploration of faculty perspectives on the in-training evaluation of residents. *Academic Medicine*, 85(7), 1157–62. <https://doi.org/10.1097/ACM.0b013e3181e19722>. PMID: 20592512.

- Watling, C., Driessen, E., van der Vleuten, C. P., Vanstone, M. and Lingard, L. (2012). Understanding responses to feedback: The potential and limitations of regulatory focus theory. *Medical Education*, 46(6), 593–603. <https://doi.org/10.1111/j.1365-2923.2012.04209.x>. PMID: 22626051.
- Weller, J. M., Jolly, B., Misur, M. P., Merry, A. F., Jones, A., Crossley, J. M., Pedersen, K., Smith, K. (2009a). Mini-clinical evaluation exercise in anaesthesia training. *British Journal of Anaesthesia*, 102(5), 633–41. <https://doi.org/10.1093/bja/aep055>. PMID: 19336537.
- Weller, J. M., Jones, A., Merry, A.F., Jolly, B. and Saunders, D. (2009b). Investigation of trainee and specialist reactions to the mini-clinical evaluation exercise in anaesthesia: implications for implementation. *British Journal of Anaesthesia*, 103(4), 524–30. <https://doi.org/10.1093/bja/aep211>. PMID: 19687033.
- Whitehouse, A., Hassell, A., Bullock, A., Wood, L. and Wall, D. (2007). 360 degree assessment (multisource feedback) of UK trainee doctors: field testing of team assessment of behaviours (TAB). *Medical Teacher*, 29(2-3), 171–76. <https://doi.org/10.1080/01421590701302951>. PMID: 17701629.
- Wilkinson, J. R., Crossley, J. G. M., Wragg, A., Mills, P., Cowan, G. and Wade, W. (2008). Implementing workplace-based assessment across the medical specialties in the United Kingdom. *Medical Education*, 42(4), 364–73. <https://doi.org/10.1111/j.1365-2923.2008.03010.x>. PMID: 18338989.
- Wilson M. and Scalise K. (2015) Assessment of Learning in Digital Networks. In: Griffin P., Care, E. (Eds.) *Assessment and Teaching of 21st Century Skills. Educational Assessment in an Information Age*. Springer, Dordrecht. p. 57–81. https://doi.org/10.1007/978-94-017-9395-7_3.
- Wragg, A., Wade, W., Fuller, G., Cowan, G. and Mills, P. (2003). Assessing the performance of specialist registrars. *Clinical Medicine*, 3(2), 131–34. <https://doi.org/10.7861/clinmedicine.3-2-131>. PMID: 12737369, PMCID: PMC4952733.
- Yeates, P., O'Neill, P., Mann, K. and Eva, K. (2013). Seeing the same thing differently. *Advances in Health Sciences Education: Theory and Practice*, 18(3), 325–41. <https://doi.org/10.1007/s10459-012-9372-1>. PMID: 22581567.
- Zendejas, B., Brydges, R., Wang, A. T. and Cook, D. A. (2013). Patient outcomes in simulation-based medical education: a systematic review. *Journal of General Internal Medicine*, 28(8), 1078–89. <https://doi.org/10.1007/s11606-012-2264-5>. PMID: 23595919, PMCID: PMC3710391.