Implementation and Use of Workplace-Based Assessment in Clinical Learning Environments: A Scoping Review

Hannah L. Anderson, MBA, Joshua Kurtz, MD, and Daniel C. West, MD

Abstract

Purpose
Workplace-based assessment (WBA) serves a critical role in supporting competency-based medical education (CBME) by providing assessment data to inform competency decisions and support learning. Many WBA systems have been developed, but little is known about how to effectively implement WBA. Filling this gap is important for creating suitable and beneficial assessment processes that support large-scale use of CBME. As a step toward filling this gap, the authors describe what is known about WBA implementation and use to identify knowledge gaps and future directions.

Method
The authors used Arksey and O’Malley’s 6-stage scoping review framework to conduct the review, including: (1) identifying the research question; (2) identifying relevant studies; (3) study selection; (4) charting the data; (5) collating, summarizing, and reporting the results; and (6) consulting with relevant stakeholders.

Results
In 2019–2020, the authors searched and screened 726 papers for eligibility using defined inclusion and exclusion criteria. One hundred sixty-three met inclusion criteria. The authors identified 5 themes in their analysis: (1) Many WBA tools and programs have been implemented, and barriers are common across fields and specialties; (2) Theoretical perspectives emphasize the need for data-driven implementation strategies; (3) User perceptions of WBA vary and are often dependent on implementation factors; (4) Technology solutions could provide useful tools to support WBA; and (5) Many areas of future research and innovation remain.

Conclusions
Knowledge of WBA as an implemented practice to support CBME remains constrained. To remove these constraints, future research should aim to generate generalizable knowledge on WBA implementation and use, address implementation factors, and investigate remaining knowledge gaps.

Medical education throughout the world has been moving to a competency-based medical education (CBME) system because of its potential to produce physicians better prepared to meet the needs of society. CBME advocates for time-variable advancement, where learners advance through training when there is sufficient evidence that they have met a fixed set of competency standards or outcomes. CBME depends on a rigorous system of assessment to provide evidence that learners have achieved competency standards. Assessments to support CBME often fall under the category of workplace-based assessment (WBA), which consists of methods to gather “evidence of clinical competence and professional behavior observed in clinical environments.” WBA serves a critical role in programs of assessment in CBME because these assessments are authentic to the day-to-day activities that a physician must perform in delivering patient care. WBA can also serve an important role in providing feedback to learners and supporting learning.

In the United States, the search for WBA processes to support CBME began in the early 2000s when the Accreditation Council for Graduate Medical Education (ACGME) and the American Board of Medical Specialties (ABMS) adopted 6 general domains in which all physicians should be competent. In 2010, Holmboe et al identified the continuing need for rigorous, continuous assessment systems rooted in WBA. Over the same time frame, a number of competency frameworks have been developed throughout the world (e.g., ACGME Milestones, CanMEDS, Entrustable Professional Activities [EPAs]), and others along with many individual assessment tools and processes to support those frameworks.

Over the past decade, research in WBA has focused primarily on building validity evidence for the use of assessment tools within competency-based frameworks and understanding how trainees and assessors experience assessment. Recent reviews of WBA tools and user perceptions have noted that use and implementation of assessments remains understudied and while WBA continues to be widely adopted, approaches to implementation and using WBA can vary widely between institutions and within the various training programs and clinical learning environments inside institutions. Importantly, research focusing on this variability in the application and implementation of WBA assessment to support CBME has been limited. Filling this research gap is important because without a clear understanding of how WBA is actually implemented and used, it is difficult to effectively create suitable, beneficial assessment processes to help realize the full potential of CBME.
The purpose of our review was to take a step toward filling this research gap by describing what is known about how WBA has been implemented and used in clinical learning environments and to identify gaps and directions for future research that could facilitate the widespread implementation of WBA to support CBME. Because we aimed to describe a large body of assessment literature from the past eighteen-plus years, we determined that a scoping review would be the best review method to achieve our purpose. Consistent with expert guidance on conducting scoping reviews, this review did not compare the relative strengths or rigor of included studies. Our review adds to the literature by synthesizing what is currently known about WBA as an implemented practice and identifying future directions for research and innovation.

**Method**

We used Arksey and O’Malley’s 6-stage scoping review framework to conduct our review, including the optional final stage of consulting stakeholders. 18,19

**Stage 1: Identifying the research question**

Our aim was to understand what is known about how WBAs have been implemented and used, both as individual WBA tools and as larger programs of assessment. Specifically, our research question was: What is currently known about WBA in implemented practice and/or use of a WBA tool or program in a clinical environment? After removing duplicates, 3 reviewers (H.L.A., D.C.W., and research assistant) applied the established inclusion and exclusion to the titles and abstracts of a randomly selected 10% sample of identified studies and calculated our agreement (Kappa = 0.86). We resolved discordance through discussion among reviewers. Because concordance exceeded our target threshold (Kappa > 0.79), the same 3 reviewers performed inclusion/exclusion classifications by discussion among reviewers. Because concordance exceeded our target threshold (Kappa > 0.79), the same 3 reviewers performed inclusion/exclusion classifications of the remaining articles independently. Two reviewers (H.L.A., D.C.W.) performed a second round of inclusion/exclusion classification using the full text of all articles that met inclusion criteria in the first screening. We used the PRISMA Extension for Scoping Reviews (PRISMA-ScR) checklist to document searches and screening. 20

**Stage 2: Identifying relevant studies**

Our research team worked with a medical librarian to develop our literature search strategy. We used combinations of the following search terms: assessment, workplace-based assessment, clinical learner assessment, clinical learning environment, clinical learning, WBA*, use, usability*, utilit*.* We searched 4 databases (PubMed, CINAHL, EBSCO, and Scopus) in September 2019 and again in March 2020 to find the most updated articles and to ensure that we included gray literature. In addition, we hand-searched the reference lists of included articles until no new article results appeared.

**Stage 3: Study selection**

To meet inclusion criteria, studies had to focus on the use and/or implementation of a WBA tool, framework, or program to assess learners in a clinical learning environment. Learners and learning environment could be from any clinical field, including medicine, dentistry, nursing, veterinary medicine, or interprofessional education. We excluded studies if: (1) a full text was not available in English, (2) it was published pre-2002 (to identify articles published subsequent to the ABMS's definition of the 6 competency domains), or (3) if the study focused on learning that did not occur in clinical workplace environments (i.e., online learning, classroom learning, or simulation environments). Additionally, we excluded studies that focused on WBA but did not explicitly consider implementation and/or use of a WBA tool or program in a clinical environment. After removing duplicates, 3 reviewers (H.L.A., D.C.W., and research assistant) applied the established inclusion and exclusion to the titles and abstracts of a randomly selected 10% sample of identified studies and calculated our agreement (Kappa = 0.86). We resolved discordance through discussion among reviewers. Because concordance exceeded our target threshold (Kappa > 0.79), the same 3 reviewers performed inclusion/exclusion classifications of the remaining articles independently. Two reviewers (H.L.A., D.C.W.) performed a second round of inclusion/exclusion classification using the full text of all articles that met inclusion criteria in the first screening. We used the PRISMA Extension for Scoping Reviews (PRISMA-ScR) checklist to document searches and screening. 20

**Stage 4: Charting the data**

We developed a data charting form to record information from the studies and address key components of the research question. (See Supplemental Digital Appendix 1 at http://links.lww.com/ACADMED/B172 for link to complete data charting form.) The initial form included the following elements: title, first author name, field of study (e.g., medicine, dental, nursing, etc.), study design, research question, methodology, and findings. Two investigators (H.L.A., D.C.W.) independently pilot tested the form on 5 included studies, after which 2 additional fields were added to the form—study limitations and directions for future research. Three investigators (H.L.A., J.K., D.C.W.) then independently charted data from each included study.

**Stage 5: Collating, summarizing, and reporting the results**

We used Microsoft Excel 2016 (Microsoft Corporation, Redmond, Washington) for descriptive analysis of publication year, clinical field, learner type, and study design of all included studies. We used thematic analysis to understand themes in the data collected in stage 4. 21 We developed initial codes through discussion with all authors. One author (H.L.A.) and a research assistant iteratively coded the data with frequent discussion and construction of codes with the entire research team. We iteratively reviewed data with related codes and categories during regular meetings until consensus on emerging themes was reached. Finally, we discussed and summarized themes and their relationships. We used Atlas.ti version 8 (Scientific Software Development, Berlin) to collect the qualitative data in a central location and organize codes.

**Stage 6: Consultation exercise**

We performed 2 consultation exercises with a group of stakeholders in medical education from our local institution and 2 outside institutions. The stakeholder group included program directors in GME, clinical educators, trainees, and education researchers. At the first consultation in December 2019, we presented our research question and goals and collected stakeholder input in developing our inclusion and exclusion criteria. During that consultation, stakeholders made specific recommendations to include only studies that discussed the clinical workplace learning environment and exclude simulation-based assessments. At the second consultation in November 2020, we asked stakeholders to read our manuscript draft and provide feedback on the usefulness of our findings for their clinical and educational practice. Stakeholders provided recommendations for writing, formatting, and presenting our results.

**Results**

**Overview**

We screened a total of 726 papers for eligibility using our established inclusion
and exclusion criteria. Of those, 163 met the inclusion criteria and were included in our review (Figure 1). Included papers were primarily from medical education (n = 139) with the remainder of papers coming from interprofessional education (n = 15), dental education (n = 4), nursing education (n = 3), and veterinary education (n = 2). Within medical education, papers focused heavily on graduate medical education (GME) (n = 76). Some papers used or discussed competency frameworks, notably ACGME Milestones (n = 21) and EPAs (n = 21). We report summary demographics of included articles in Table 1 and the full list of included papers in Supplemental Digital Appendix 2, available at http://links.lww.com/ACADMED/B172.

We identified 5 themes in our thematic analysis of the literature:

1. Many WBA tools and programs have been implemented, and barriers to successful implementation are common across fields and specialties.
2. Theoretical perspectives consistently emphasize the need for data-driven implementation strategies.
3. User perceptions of WBA vary and are often dependent on implementation factors.
4. Technology solutions could provide useful tools to support WBA.
5. Many areas for future research and innovation remain.

**Theme 1: Many WBA tools and programs have been implemented, and barriers to successful implementation are common across fields and specialties**

The implementation and use of a single assessment tool to assess specific skills (e.g., competency in a clinical or procedural task) in specific types of clinical learning environments (e.g., a single clinical unit or department) was the most commonly implemented form of WBA found in our review, with a total of 31 pilot studies. While individual WBA tools were the most commonly studied, 26 papers discussed implementing pilots of programmatic assessment, where multiple individual assessment tools are used by multiple assessors in different contexts to

---

**Figure 1** PRISMA Extension for Scoping Reviews (PRISMA-ScR) exclusion, inclusion, and review process for 726 manuscripts focusing on implementation of workplace-based assessment, 2019–2020.
support an educational design.

Since 2005 when programmatic assessment was defined, programs of assessment primarily have been built to support undergraduate medical education (UME) curriculums, although some studies reported its use in GME contexts. We identified several salient themes on implementation from studies of programmatic assessment, including: (1) the challenge of adapting to lack of resources and variability of resources in different learning environments, (2) the necessity of implementing complex assessment programs iteratively, and (3) the difficulty of meaningfully evaluating program outcomes. Studies of programmatic assessment were typically implemented at a single institution and reported 1–4 years of outcome data.

Studies investigated aspects of validity most commonly, with the majority of studies concluding that the relevant tool was valid for its intended purpose.8,9,22,25,26,30,33,36,37,39,41,46,48,50,53–75 Eight studies investigated the use and implementation of multisource feedback tools, which were the most prevalent type of implemented tool.6,54,55,59,60,62,63,68

Several papers emphasized the importance of including the viewpoints of multiple assessors in the clinical learning environment and the importance of collecting both quantitative and qualitative data in assessment tools.70,76–87

When discussing the analysis and interpretation of qualitative comments that assessors used to describe trainee competency, qualitative comments were not only valid and useful for drawing conclusions about competence but also considered an essential element of WBA.55,64,66,80

In our thematic analysis, we identified barriers to successfully design and implement WBA tools that were reported in UME and multiple specialties in GME. Common barriers included: (1) lack of trainee and assessor engagement in design, (2) time constraints of the clinical environment, (3) and distilling the complex language of competency-based assessment into terms and parameters that assessors and trainees could easily use.

**Theme 2: Theoretical perspectives emphasize the need for data-driven implementation strategies**

Over the time frame of our review, the purpose of conceptual papers shifted from proposing and developing frameworks for designing assessment to frameworks and strategies for implementing and using assessment. Several perspectives, commentaries, and framework proposals discussed concepts for developing an overall strategy for implementation of WBA as part of programmatic assessment.4,5,7,65,88–115

In alignment with programmatic assessments’ ideals, a key element of these frameworks was the use of multiple assessment points to inform summative decisions.100–102,116–119 Recommendations included combining and using assessment data from multiple contexts and different assessors, paying close attention that assessments have utility in their particular learning environment.120,121 One study emphasized the importance of collecting and displaying large amounts of data systematically and reliably.107 However, papers rarely recommended practical or technical ways to accomplish that task besides the use of assessment portfolios to gather evidence. For example, one paper suggested that programs of assessment may benefit from the use of assessment portfolios in continuing medical education (CME).111

Another important implementation strategy was prioritizing embedding WBA into preexisting clinical activities and workflows; however, this was deemed challenging to achieve while preserving authentic use of assessments.7,9,122 Designing assessment systems that accurately support competency frameworks without burdening trainees and assessors with complex, unfamiliar language and concepts was reported as a challenge.81,110,123 Approaches for overcoming these challenges included ensuring evaluation forms used easy-to-understand language, investing time and effort into assessor training, and developing assessors’ shared mental frameworks for learner assessment.4,7,124,125

**Theme 3: User perceptions of WBA vary and are often dependent on implementation factors**

Almost all studies on user perceptions were limited in scope to a single institution and/or single specialty, thus limiting the ability to generalize these findings. Nevertheless, one consistently identified important implementation strategy was the need to obtain assessor and trainee buy-in. A commonly reported key to achieving buy-in was designing and implementing specific WBA tools or programs so that they were meaningful to users. For trainees, meaning was enhanced when WBA processes were seen as having educational value; conversely, trainees found less value in WBA when learning and educational
A concern specific to assessors was that WBA tools and systems required a significant amount of their time and effort. For example, significant assessor training in how to perform the assessment and complete the tool was a critical component in making WBA acceptable, feasible, and valid. However, the assessor time required to engage in training was also a barrier to successful implementation. Three papers investigated user concerns about establishing trust and reducing assessor bias and/or subjectivity. Some studies mentioned assessor training as a feasible way to reduce impacts of assessor bias. Four papers reported trainees concerns about equity and unfairness in the learning environment but did not describe features of implementation or use that contributed to those perceptions.

Theme 4: Technology solutions could provide useful tools to support WBA

Early WBA studies used paper-based forms, including feedback forms, to document, record, and share assessment data in local contexts. Following the move to the ACGME Next Accreditation System and the requirement for reporting ACGME Milestone ratings in 2013, studies shifted focus to implementing computer- and web-based tools to meet increasing demands for data usage and storage. Surgical specialties in particular reported using multiple platforms to upload and store data reflecting a range of technological solutions, including web and mobile applications. These platforms were designed to meet the needs of busy surgical clinical environments where assessments are often difficult to complete under time constraints. Additionally, platforms used in surgical specialties were often designed and implemented with direct input from trainees and assessors, which was reported as a beneficial method for creating acceptable, usable technology tools. In general, technology solutions were well received, but studies noted that these solutions currently have limited outcomes data and their long-term effectiveness has not been determined.

E-portfolios have been used to support assessment systems, especially to support providing feedback to learners. Successful use of E-portfolios was reported to depend on effective infrastructure and program support. In recent years, studies reflected an increasing interest in learning analytics to make assessments more accessible and useful for trainees, assessors, program leadership, and Clinical Competency Committees (CCC).

Theme 5: Many areas of future research and innovation remain

Studies across specialties and interprofessional education recommended further investigation of WBA as a way to extend CBME into continuing professional development (CPD) (e.g., CME, maintenance of certification, etc.). Studies on programmatic assessment called for further studies on WBAs relationship to educational outcomes and how WBAs that are used for both formative and summative purposes influence learning behaviors and assessor behaviors. Other studies called for further investigation into the role and impacts of assessor subjectivity, bias, and/or variability especially when WBA is used to inform summative decisions.

Many studies identified the need for more effective technology to support WBA programs as an important area for future research. Examples include the need to develop databases for data reporting and synthesis to provide feedback to learners and support summative decision making and develop better understanding of how to use E-portfolios and other methods to track assessment data across the UME/GME continuum. Finally, studies consistently identified the necessity of rigorously implementing WBA on a larger scale (i.e., across multiple specialties or institutions) to collect larger sets of data and study broader educational, clinical, and cost outcomes.

Discussion

We identified 5 themes that describe what is currently known about the usage and implementation of WBA in clinical learning environments. Central to the first theme, we found that many WBA tools, and some programs of assessment, have been implemented, and studies of individual WBA tools reported evidence supporting the validity of using tools for their intended purpose. However, the generalizability of the findings from these studies is limited because they were usually implemented in single institutions with a single-specialty focus and often reported only short-term outcomes data. Furthermore, many of the theoretical papers we reviewed recommended considering local context, stakeholders, and other barriers and facilitators in the design and implementation of WBA, yet we found that few studies addressed these issues in any systematic way. Our findings across each of the 5 themes suggest 3 key core directions for future research and innovation (Figure 2) that the field should pursue to move the field of WBA forward as an implemented practice to support the use of CBME on a large scale for its intended purpose.

First, future work should aim to generate greater generalizable knowledge on WBA through large-scale, longitudinal implementations across multiple institutions and/or specialties. Large-scale implementation studies could fill critical gaps in our knowledge of how to effectively and efficiently implement WBA. In addition, more longitudinal studies of both small- and large-scale implementations (i.e., longer than 1- to 4-year pilot programs) could help develop a deeper understanding of WBAs’ usefulness, meaning, and impacts on users. Similar to approaches that have been used to promote the uptake of evidence-based practices in patient care, the design of future studies should include implementation science and human factors research methods. The use of these methodologies has the potential to greatly advance our understanding of how to effectively implement WBA programs and advance the uptake of practices that have been deemed important in theoretical and small-scale work.
Second, key barriers to implementation of WBA that are already known must be addressed. In our review, we found that reported barriers to successful implementation and effective use of WBA were numerous. However, 3 barriers were commonly reported and overlapped across 3 of our identified themes (themes 1, 2, and 3). These common barriers included: (1) user concerns about the impact of WBA tools and programs on educational experiences and learning; (2) the burden of WBA on assessors and local resources; and (3) the challenge of implementing WBA in ways that effectively integrate into, rather than interfere with, activities and workflows in the clinical learning environment. Despite the limitations in generalizability we discussed previously, these barriers were so commonly reported that developing methods to overcome them, such as securing user buy-in and input during assessment design and implementation rather than postdevelopment, seems to be a vital next step.

The potential for technology solutions to support the successful implementation of WBA was important theme that we identified in our review, and it represents one area where elements of implementation science and human factors research have been used effectively. For example, concerns about limited time to complete assessments in surgical specialties frequently drove the development and implementation of user-friendly, streamlined online assessment applications. These tools were often designed by direct user input and tailored to the needs of the local environment. These examples are an encouraging demonstration that when implementation barriers are addressed, WBA can be acceptable to users, even in time- and resource-constrained learning environments. Outside of surgical specialties, however, we found little evidence of the development of technology solutions designed to efficiently collect assessment data in the workplace. Furthermore, once WBA assessment data are collected, the ability to store, retrieve, display, and otherwise “make sense” of these data is vital to support learning and making advancement decisions in CBME. Future research should focus on these gaps by investigating new technologies, adapting previously studied technologies for use in new clinical learning environments, and creating platforms that synthesize assessment data for stakeholder use.

Finally, a broad range of other significant gaps in our understanding of WBA must be filled. Several theoretical perspectives papers noted the need to study the relationship between WBA and learning behaviors, such as mastery learning, master-adaptive learning, and self-directed learning; however, we found that only 2 studies included in this review explicitly investigated learning behaviors. While these studies collected data on the learning behaviors of trainees such as feedback seeking and receptivity within the context of a program of assessment, they did not report longitudinal data on those features. As a result, understanding the relationship between WBA and broader, longer-term educational outcomes and learning behaviors is an important gap that needs to be filled. Additionally, while WBA in CPD has been considered useful theoretically, especially as a tool to support maintenance of certification activities of physicians, actual implementation of WBA across the continuum of CBME into CPD contexts deserves much more study.

We found that a limited number of studies addressed the role of subjectivity and bias in WBA at the level of individual assessors. However, no studies explicitly considered or measured other aspects of WBA that might impact equity and inclusion in clinical learning environments. Important gaps include trainee perceptions of subjectivity and/or bias in high-stakes assessments intended to inform remediation, advancement,
Our review does have some limitations. We only included papers that had full texts available in English; therefore, we could have missed relevant papers in other languages. It is also possible that we overlooked some relevant papers during the literature search process. However, due to the large number of included papers and our hand-searching of reference lists and gray literature, we feel confident that we included appropriate papers that represent the literature related to implementation of WBA.

In this review, we provide a thematic description of our current understanding of WBA as an implemented practice and identify a number of important gaps in our knowledge. While WBA tools, programs, technology systems, and theoretical perspectives have been developed across many clinical fields and specialties, our knowledge of WBA as an implemented practice remains constrained. Removing these constraints is an important next step to facilitate the widespread implementation of CBME to achieve its intended purpose. Future research should aim to generate generalizable knowledge on WBA implementation and use, address key implementation determinants, and investigate key remaining gaps in our knowledge.

Acknowledgments: The authors thank the members of The Children’s Hospital of Philadelphia Education Collaboratory for their valuable input on this review during its design and preparation, and Alyssa Sze, research assistant, for her assistance during the data charting and collating stages.

Funding/Support: None reported.

Other disclosures: None reported.

Ethical approval: Reported as not applicable.

References
Workplace Assessment


60 Lee AG, Oettting T, Beaver HA, Carter K; Task Force on the ACGME Competencies at the University of Iowa Department of Ophthalmology. The ACGME outcome project in ophthalmology: Practical recommendations for overcoming the barriers to local implementation of the national mandate. Surv Ophthalmol. 2016;61:469–481.


Hsiao CT, Chou FC, Hsieh CC, Chang LC, Edwards J, Petra H. The effects of external
Donato AA, George DL. A blueprint for
Bok HG, Teunissen PW, Favier RP, et al. Learning and assessment system for residency
Hsu CM. Developing a competency-based
Miller A, Archer J. Impact of workplace
Dijkstra J, Van der Vleuten CP, Schwurich LW. A new framework for designing
ten Cate O. Entrustment assessment: Recognizing the ability, the right, and the duty to act. J Grad Med Educ. 2016;8:261–262.
Kalsi HK, Kalsi JS, Fisher NL. An
Jones LK Jr. Assessment in neurologic
121 Kalsi HK, Kalsi JS, Fisher NL. An
112 ten Cate O. Entrustment assessment: Recognizing the ability, the right, and the duty to act. J Grad Med Educ. 2016;8:261–262.


