

Workplace-Based Assessments Using Pediatric Critical Care Entrustable Professional Activities

Amanda R. Emke, MD, MHPE

Yoon Soo Park, PhD

Sushant Srinivasan, MD

Ara Tekian, PhD, MHPE

ABSTRACT

Background Workplace-based assessment (WBA) is critical to graduating competent physicians. Developing assessment tools that combine the needs of faculty, trainees, and governing bodies is challenging but imperative. Entrustable professional activities (EPAs) are emerging as a clinically oriented framework for trainee assessment.

Objective We sought to develop an EPA-based WBA tool for pediatric critical care medicine (PCCM) fellows. The goals of the tool were to promote learning through benchmarking and tracking entrustment.

Methods A single PCCM EPA was iteratively subdivided into observable practice activities (OPAs) based on national and local data. Using a mixed-methods approach following van der Vleuten's conceptual model for assessment tool utility and Messick's unified validity framework, we sought validity evidence for acceptability, content, internal structure, relation to other variables, response process, and consequences.

Results Evidence was gathered after 1 year of use. Items for assessment were based on correlation between the number of times each item was assessed and the frequency professional activity occurred. Phi-coefficient reliability was 0.65. Narrative comments demonstrated all factors influencing trust, identified by current literature, were cited when determining level of entrustment granted. Mean entrustment levels increased significantly between fellow training years ($P = .001$). Compliance for once- and twice-weekly tool completion was 50% and 100%, respectively. Average time spent completing the assessment was less than 5 minutes.

Conclusions Using an EPA-OPA framework, we demonstrated utility and validity evidence supporting the tool's outcomes. In addition, narrative comments about entrustment decisions provide important insights for the training program to improve individual fellow advancement toward autonomy.

Introduction

Workplace-based assessment (WBA), the process of observing and assessing trainee performance in the clinical environment, is critical to the development and certification of competent physicians. Directly observing trainees in the workplace provides opportunities for immediate performance feedback to guide further learning.¹ By situating the learning within daily clinical duties and focusing on assessment that improves performance, WBAs present the greatest rewards for all stakeholders.²⁻⁴

The challenge has been identifying a WBA framework that balances holistic and longitudinal judgments of trainee performance with appropriate granularity for frontline clinicians. Based on the requirement to upload milestone levels for trainees every 6 months, the Accreditation Council for Graduate Medical Education (ACGME) Competencies and Milestones served as an initial framework for

WBA tools.⁵⁻⁸ Unfortunately, many behaviors in the Pediatrics Milestones, such as “reflection in” versus “reflection on” action, cannot be directly observed, and tools based on the ACGME Competencies and Milestones deconstructed the complex interactions needed to make clinical decisions. This mismatch between the assessments by clinical faculty and the way clinicians function in practice leads to construct misalignment which creates unreliable assessment outcomes.⁹

Entrustable professional activities (EPAs), descriptions of daily professional activities trainees should be capable of performing independently, emerged as an alternative WBA framework.¹⁰ EPAs provide a holistic approach to defining workplace outcomes and having improved construct alignment.¹⁰⁻¹² While not required by the ACGME as part of trainee assessment, EPAs developed by specialty societies have provided a mechanism for converting theoretical concepts into meaningful clinical curricula and WBA tools by standardizing language for the decisions made by clinical faculty about trusting and supervising trainees.¹⁰⁻¹⁵ The scope of some EPAs impacts their usefulness as an assessment for learning tool, because faculty feel uncomfortable making decisions

DOI: <http://dx.doi.org/10.4300/JGME-D-18-01006.1>

Editor's Note: The online version of this article contains full details on tool development, faculty and fellow training, and implementation of the workplace-based assessment tool.

of entrustment or supervision over such a broad statement.¹³ To address this concern, Warm and colleagues introduced the concept of observable practice activities (OPAs).^{14,15} Narrowing the scope of broad EPAs to specific practice activities provides the granularity and contextualization needed for direct observation, formative assessments, and translation to a useful assessment tool. As such, using a combined EPA-OPA framework to assess more granular, yet frequent, measurements may create useful WBA tools to ensure independent practice.

To address a paucity of meaningful assessment data for performance improvement and advancement decisions due to construct misalignment coming from current assessment tools based on the ACGME Competencies and Milestones in our PCCM fellowship, we sought to develop a novel WBA tool using an EPA-OPA framework. This study examines the development and implementation of that tool, hypothesizing that the tool provides sufficient validity evidence to infer utility as a formative assessment.

Methods

Setting and Participants

Participants included 15 faculty supervisors and 12 PCCM fellows practicing in a 28-bed medical-surgical pediatric intensive care unit at an academic, quaternary freestanding children's hospital. Data were gathered over a single academic year (2015–2016). The clinical service model paired a fellow and a faculty member together as a clinical care team for 7 days.

Tool Development

To ensure the tool overcomes current assessment barriers and meets the needs of users, PCCM fellows and faculty participated in surveys, focus groups, and individual discussions. These identified perceptions of current assessment tools and guided tool development by articulating goals for the assessment tool. Needs assessment outcomes demonstrated the previous tool's length and complex, educationally-based language as key deterrents to timely completion and provision of narrative feedback, an important step in using the assessments to enhance learning. To create the assessment items, the PCCM EPA, "acute management of the critically ill patient, including those with underlying chronic disease," served as the foundation for the WBA tool (FIGURE 1).¹⁶ Initial meetings with faculty discussing tool development revealed unwillingness to provide entrustment decisions, because the scope of this EPA was too broad. In considering the literature defining EPAs, including number and breadth, and van der Vleuten's definition

What was known and gap

Developing assessment tools that combine the needs of faculty, trainees, and governing bodies is challenging but imperative.

What is new

A novel WBA tool using a combined entrustable professional activity and observable practice activity (EPA-OPA) framework.

Limitations

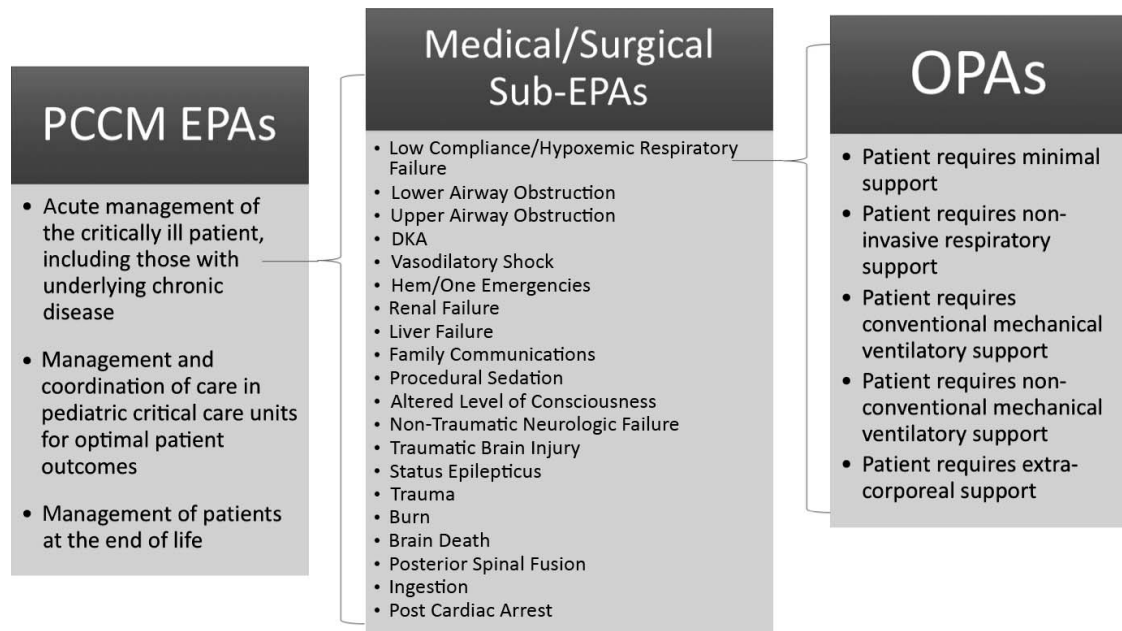
Study was conducted at a single center with a small amount of faculty and fellows, limiting generalizability. Only 1 year of data using the tool is available.

Bottom line

The EPA-OPA tool demonstrated sufficient validity evidence to justify using the outcomes to guide assessment for learning and promote the development of entrustment.

of utility, the PCCM EPA was subsequently divided into a list of sub-EPAs and OPAs.^{13,17} Creation of the initial list of sub-EPAs, based on patient management tasks, utilized the American Board of Pediatrics PCCM subspecialty examination content, the opinions of PCCM faculty with more than 5 years of experience, and the frequency of admission pathophysiology determined by sampling 3 months of admission diagnoses (FIGURE 1). An acceptable level of granularity was still unmet for faculty agreement to assess based on the sub-EPAs. A third round of refinement provided OPAs based on more granular management tasks within each sub-EPA, utilizing the mechanism described by Warm and colleagues (FIGURE 1).¹⁴ These were defined as sub-EPAs and OPAs rather than specifications or nested activities from larger EPAs, because they continued to constitute a distinct and independent unit of work that a fellow must be entrusted to perform and, within themselves, included several specifications such as interpreting vital signs and laboratory data, managing patient-assistive technology, and recognizing when to escalate or deescalate therapies.¹³ The final tool included 20 sub-EPAs with individualized OPAs. Full details on tool development, faculty and fellow training, and implementation are available as online supplemental material.

To disseminate the tool, a password-protected, conditional branching SurveyMonkey form was developed, pilot tested by faculty and fellows, and downloaded to division-provided cell phones. Tool completion required selection of the EPA-OPA (jointly chosen by faculty-fellow dyads) and faculty members to provide the level of entrustment (TABLE) and a narrative explanation for the choice and requirements for advancement to the next level. Finally, 90-minute sessions were held separately for faculty and fellows to introduce the new tool.

**FIGURE 1**

Development of Pediatric Critical Care Medicine Entrustable Professional Activity–Based Assessment Tool

Data Analysis for Validity Evidence

A mixed-methods analysis was conducted using sources of evidence outlined in Messick's validity framework (content, response process, internal structure, relations to other variables evidence, and consequences) in addition to acceptability as described in van der Vleuten's model of assessment tool utility.^{17–19} Participant demographic characteristics were obtained. Acceptability evaluation included analysis of feasibility and compliance. Time required for tool completion, the main complaint for previous assessment tools, served as the feasibility evaluation. Compliance was determined, for faculty and fellows, using each person's expected number of completed assessments based on service requirements. Because WBAs provide unequal numbers of observations per trainee from different evaluators, reliability was estimated using an unbalanced random-effects model of generalizability theory with EPAs crossed with learners, where learners were the object of measurement, and variance components of the assessment looking for phi-coefficient > 0.3 and a person variance $> 5\%$, respectively.^{20–22} Analysis of variance was used to determine the significance between mean entrustment levels using SPSS Statistics 23 (IBM Corp, Armonk, NY).

A qualitative analysis of faculty-identified rationale behind entrustment decisions was conducted using iterative thematic coding processes. A hybrid deductive and inductive approach began with 5 factors influencing entrustment decisions identified in the

literature.^{23,24} Themes corresponding to these factors as well as inductively and iteratively identified subthemes were coded in faculty narrative comments describing the rationale behind the entrustment level assessed. Narrative comments were independently coded with discussion and refinement of differences in coding. Coding was iterative, allowing for review and recoding of previous narratives until full agreement was reached. The analysis was completed by 2 authors, 1 from the implementation institution (A.R.E.) and 1 from outside the institution (S.S.).

The study was approved by the Institutional Review Boards of Washington University School of Medicine and the University of Illinois at Chicago.

Results

Demographic Characteristics

Of the 12 fellows included in the analysis, 5 (42%) were first-year fellows, 3 (25%) were second-year fellows, and 4 (33%) were third-year fellows. Of the 15 faculty, 4 (27%) were instructors, 10 (67%) were assistant professors, and 1 (7%) was an associate professor. Faculty were 60% (9 of 15) male and fellows were 42% (5 of 12) male.

Acceptability Evidence

Acceptability was determined using feasibility, based on the time required to complete the tool and compliance with twice-weekly assessment tool completion. On average, full completion of the tool

TABLE

Levels of Entrustment and Definitions Provided to Faculty and Fellows

Entrustment Level	Definition
Observation	▪ Trainee observes decision making and patient care; faculty make or direct all management decisions ▪ Fellow is at this level if faculty make all the decisions, even if faculty allow the fellow to enact decisions
Direct supervision	▪ Proactive supervision ▪ Faculty are standing by the fellow discussing every decision ▪ Fellow may make some or minor clinical decisions, but is not allowed full independent decision making
Indirect supervision	▪ Fellow is making most or major clinical decisions or managing a patient or situation with faculty quickly available on request ▪ Faculty are present in the unit for questions or immediate assistance ▪ Fellow develops a plan for patient care or management and discusses with faculty prior to enacting ▪ If faculty are standing with the fellow during patient care, faculty are allowing fellow to make all the decisions or conduct all patient care without faculty input but with debriefing after the encounter
Unsupervised	▪ Supervision of the fellow at a distance or post-hoc ▪ Fellow conducts patient care or makes medical decisions and only discuss with faculty after enacting ▪ Faculty are in their office or at home
Supervises others	▪ Fellow make clinical decisions or conducts patient care as in Level 4 (Direct supervision) by directing others (residents or advanced practice nurses) as defined in Level 2 (Unsupervised)

(initiation of the assessment to faculty signature) required 4.6 minutes, with a range of 1 to 14.3 minutes. To create dual responsibility over assessment, the tool was downloaded on division-provided cell phones for the fellows. As such, compliance was determined for individual faculty and fellows as the

percentage of assessments completed compared to those expected. Faculty were 50% to 100% compliant with twice-weekly completion. All faculty completed assessments at least once each service week. Fellows demonstrated wider variability with 28% to 100% compliance.

Content Validity Evidence

Content validity evidence sought to demonstrate that the items selected for assessment were representative.^{18,25} Twenty sub-EPAs were chosen *a priori* as the assessment items using experts. A total of 171 assessments were completed. Sub-EPA assessments ranged from 0 to 24. Only 1 received no assessments with a median of 6 assessments per sub-EPA. Of these 20 sub-EPAs, 19 (95%) were assessed at least once with 14 (70%) assessed at least 5 times. When compared to admission diagnoses, the frequency of sub-EPA assessment demonstrated similar distribution suggesting appropriate choices for assessment items. Individual OPA assessments ranged from 0 to 9 with most receiving 1 assessment.

Internal Structure Validity Evidence

Evidence for internal structure validity sought to demonstrate reliability of trainee entrustment level assigned by showing that differences in entrustment were due to the trainee, not assessors. Person variance (true variance) based on generalizability theory was 7.2% of total variance. The phi-coefficient reliability between person and score was 0.65.

Response Process Evidence

Response process validity evidence sought to demonstrate that the responses provided were accurate by understanding faculty rationale for providing a given entrustment level. Medical education literature has identified 5 factors that influence entrustment decisions: (1) trainee factors, (2) supervisor factors, (3) trainee-supervisor relationship factors, (4) task-based factors, and (5) factors regarding the context in which the entrustment decision occurs.^{26–30} These 5 factors influencing entrustment decisions served as the initial categories for thematic analysis to determine if faculty were using the same factors to assign entrustment levels.

In total, 171 faculty comments were included for analysis. Within the thematic analysis, each of the 5 previously identified trust influencers appeared, and an additional 7 subthemes were identified within “trainee” (FIGURE 2). Due to the complexity of select faculty narratives, multiple codes were applied to some comments. Descriptions of themes and

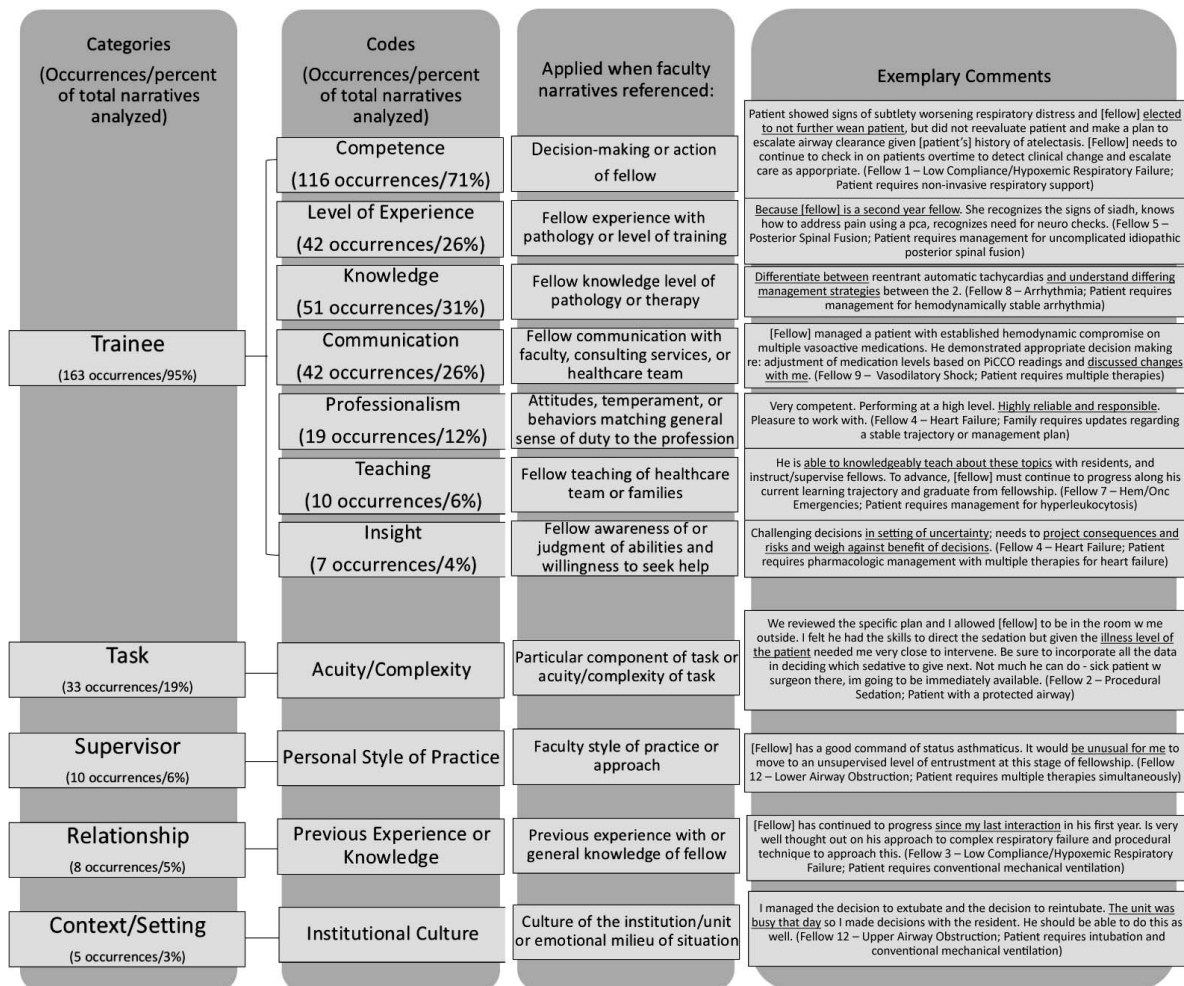


FIGURE 2

Organizational Model for Categories and Codes of Faculty Narratives Describing What Influenced Entrustment Decision

subthemes as well as illustrative comments are included in FIGURE 2. As multiple codes were often applied to comments, emphasis added by the authors identifies specific phrases or words leading to the application of the described code. “Trainee” was the most commonly applied code, seen in 163 narratives. “Task,” “supervisor,” “relationship,” and “context/setting” were applied with decreasing frequency.

Relations to Other Variables’ Validity Evidence

Relations to other variables’ evidence is derived from comparing total scores to subsets scores, correlation of novel assessment tool outcomes to a gold standard, or correlation with an expected outcome.¹⁹ Because a gold standard for WBA is absent, correlation with an expected outcome and the ability of the tool to differentiate between years of fellowship training provided evidence of relations to other variables. Mean entrustment levels (FIGURE 3) increased

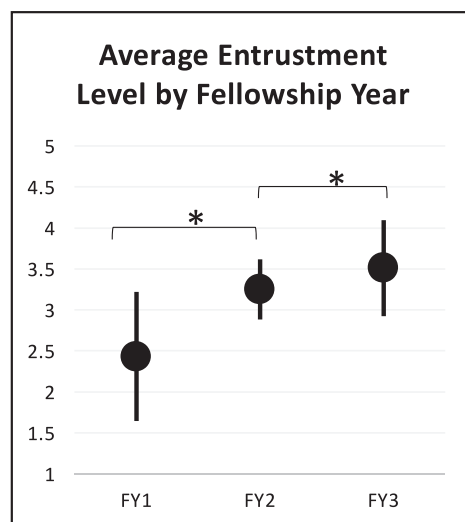


FIGURE 3

Mean Entrustment Level Across All Sub-EPAs by Year of Fellowship Training

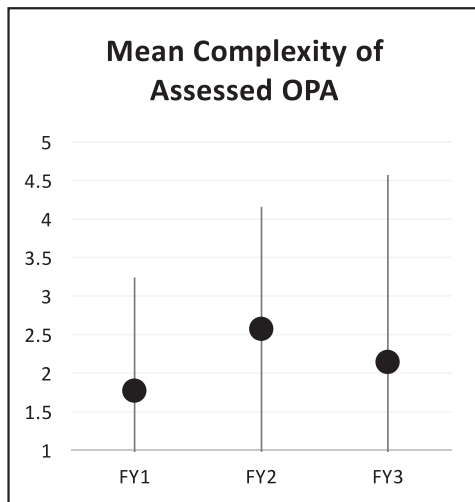


FIGURE 4
Subanalysis of Impact of Average OPA Complexity on Assigned Entrustment Level Using Most Frequently Assessed Sub-EPA Low Compliance/Hypoxemic Respiratory Failure Demonstrated No Significant Differences Between Year of Fellowship Training

significantly between training years across all EPAs ($P = .001$; fellowship year 1 [FY-1]: mean 2.43, SD 0.38; FY-2: mean 3.25, SD 0.17; FY-3: mean 3.51, SD 0.28). We also sought to understand the impact of task complexity, as this could be used by the program to ensure trainees were managing more complicated patients and/or that trainees were given more entrustment as training advanced. The sub-EPA “Low Compliance/Hypoxemic Respiratory Failure,” the most frequently assessed sub-EPA for all fellow years with 24 total assessments, was used to understand this educational impact. Analysis of this sub-EPA comparing average OPA complexity to year of training did not reveal significant differences. First-year fellows demonstrated a mean complexity of 2.77 out of 5 (SD 0.73) compared to 3.57 (SD 0.79) and 3.14 (SD 1.21) in FY-2s and FY-3s, respectively ($P = .52$; FIGURE 4). When assessed across the same complexity level, an increase in entrustment level was seen between years of fellowship training, but it did not reach statistical significance. First-year fellows demonstrated a mean entrustment level of 2.38 (SD 0.51) compared to 3.14 (SD 0.38) in FY-2s and 3.86 (SD 0.90) in FY-3s ($P = .08$).

Discussion

The novel EPA-OPA tool described in this study demonstrated appropriate validity evidence to support its use as a tool to guide assessment for learning and address some of the current concerns surrounding

WBA tools. The results of this study also identified how the tool might be used as trainees advance and the faculty-documented reasons behind supervision decisions that can help guide learning.

The content validity evidence supports that the tool measured important practice activities, a key factor needed to support learning. Our work demonstrated that family communication, while not a pathophysiology- or diagnosis-based assessment item, was one of the most frequently assessed activities. Adult studies have identified family-physician communication as a key factor in satisfaction with care provided and the psychological well-being of family members of critically ill patients,^{31–33} supporting the notion that “Family Communication” is an important practice activity for PCCM trainees developing trust to practice independently. Further support for using the outcomes of the assessment tool was demonstrated by its reliability and generalizability data. With learners as the object of measurement, sound reliability estimates and variance components estimates demonstrate the tool’s ability to discriminate between entrustable decisions. These findings, similar to those described in other generalizability studies for WBA where person variance ranges from 5% to 11%, support using the tool to discriminate between high and low performance.^{20–22}

Feasibility of a WBA tool directly relates to its length, complexity, ability to fit direct observation into clinical workflow, and the overall burden placed on faculty to complete it.^{1,2,34,35} This, in turn, influences the number of assessments completed and the ability to mitigate natural variations in performance.^{1,2,34,36} These are important threats to reliability^{18,19} that impact use of the tool for reflection and learning. The overall high number of assessments completed suggests that the tool was feasible. High response rates likely resulted from 2 decisions in tool development: (1) the choice to use the EPAs as the tool’s conceptual framework capitalizes on decisions intuitively made by clinical supervisors, and (2) the choice to involve frontline users in tool development to help decrease complexity and improve buy-in. Time constraints, especially in a busy clinical environment, significantly influence perceptions of feasibility and become detractors for completing WBA.³⁵ Considering the brief time required for completion, the evidence presented here suggests the tool should not detract from clinical workflow. The feasibility demonstrated by this tool supports development of assessment tools for the other 2 PCCM-specific EPAs: “manage and coordinate care in pediatric critical care units for optimal patient outcomes” and “management of patients at the end of life.”¹⁶

The ability to use the assessment outcomes to guide reflection and learning is also influenced by alignment with outcomes of current assessments with the same construct. Examples include comparing to a gold standard (relationship to other variables' evidence), and appropriate rationale by the person completing the assessment for the decision made (response processes evidence).^{18,25} Unfortunately, "gold standard" assessment tools in WBA are lacking especially as the ACGME Competencies and Milestones were "never intended to serve as a regular assessment tool."³⁷ However, narrative comments by faculty provided important insight into the rationale used to determine the level of entrustment afforded to a trainee. Fellows can use these narrative comments and changes in entrustment level to analyze their performance and develop a learning plan to further advance entrustment. Trainee-based factors, specifically the themes of knowledge and competence, were the most frequently cited influences of entrustment and supervision decisions reported by faculty, which mirrors the literature.^{27–29} Using these faculty narratives, fellows could develop a learning plan to read more about specific pathologies and improve their communication around decision-making. Once specific learning plans have been devised based on the assessment outcomes, fellows can use this information to guide subsequent faculty interactions, teaching, and assessments.

One unexpected finding was that, while mean entrustment levels increased by year of training, the average acuity of the practice activity assessed did not increase. One explanation is that the highest level of entrustment, "supervises others," was being provided to lower acuity patients. Decisions of entrustment and supervision are, by their nature, about risk acceptance.^{27–30} When allowing a trainee to supervise someone else, faculty may mitigate that risk by selecting lower acuity patients. Further evaluation will need to explore this idea.

This study has limitations. It was conducted at a single center with small numbers of faculty and fellows, limiting generalizability to other settings and subjects. Only 1 year of data are available for evaluation leading to cross-sectional analysis rather than longitudinal analysis in tool evaluation. The study did not determine if tool use affected trainee progress or whether trainees found the information useful. The amount of time to complete the tool and adherence rates for use of the tool are not adequate surrogate measures for tool acceptability to faculty, and additional concerns may exist.

Further studies will focus on determining how the outcomes of this assessment tool guide trainee learning, influence program changes, and can serve

as an assessment of learning by understanding how achievement of individual OPA entrustment influences sub-EPA trust decisions.

Conclusion

The EPA-OPA tool described here demonstrated sufficient validity evidence to justify using the outcomes to guide assessment for learning and promote the development of entrustment. The processes of developing, implementing, and evaluating this tool can be broadly applied.

References

1. Norcini J, Burch V. Workplace-based assessments as an educational tool: AMEE guide No. 31. *Med Teach*. 2007;29(9):855–871. doi:10.1080/01421590701775453.
2. van der Vleuten CPM, Schuwirth LW, Driessen EW, Dijkstra J, Tigelaar D, Baartman LK, et al. A model for programmatic assessment fit for purpose. *Med Teach*. 2012;34(3):205–214. doi:10.3109/0142159X.2012.652239.
3. van de Wiel MW, van den Bossche P, Janssen S, Jossberger H. Exploring deliberate practice in medicine: how do physicians learn in the workplace? *Adv Health Sci Educ Theory Pract*. 2011;16(1):81–95. doi:10.1007/s10459-010-9246-3.
4. Ericsson KA. Deliberate practice and acquisition of expert performance: a general overview. *Acad Emerg Med*. 2008;15(11):988–994. doi:10.1111/j.1553-2712.2008.00227.x.
5. Accreditation Council for Graduate Medical Education. The Pediatrics Milestone Project. http://www.acgme.org/Portals/0/PDFs/Milestones/320_PedsMilestonesProject.pdf. Accessed June 17, 2019.
6. Nasca TJ, Philibert I, Brigham T, Flynn TC. The next GME accreditation system – rationale and benefits. *N Engl J Med*. 2012;366(11):1051–1056. doi:10.1056/NEJMs1200117.
7. Schumacher DJ, Lewis KO, Burke AE, Smith ML, Schumacher JB, Pitman MA, et al. The pediatrics milestones: initial evidence for their use as learning road maps for residents. *Acad Pediatr*. 2013;13(1):40–47. doi:10.1016/j.acap.2012.09.003.
8. Pangaro L, ten Cate O. Frameworks for learner assessment in medicine: AMEE guide No. 78. *Med Teach*. 2013;35(6):e1197–e1210. doi:10.3109/0142159X.2013.788789.
9. Crossley J, Jolly B. Making sense of work-based assessments: ask the right questions, in the right way, about the right things, of the right people. *Med Educ*. 2012;46(1):28–37. doi:10.1111/j.1365-2923.2011.04166.x.

10. ten Cate O, Scheele F. Competency-based postgraduate training: can we bridge the gap between theory and clinical practice? *Acad Med.* 2007;82(6):542–547. doi:10.1097/ACM.0b013e31805559c7.
11. Carraccio C, Englander R, Holmboe ES, Kogan JR. Driving care quality: aligning trainee assessment and supervision through practical application of entrustable professional activities, competencies, and milestones. *Acad Med.* 2016;91(2):199–203. doi:10.1097/ACM.0000000000000985.
12. Rekman J, Gofton W, Dudek N, Gofton T, Hamstra SJ. Entrustability scales: outlining their usefulness for competency-based clinical assessment. *Acad Med.* 2016;91(2):186–190. doi:10.1097/ACM.0000000000001045.
13. ten Cate O, Chen HC, Hoff RG, Peters H, Bok H, van der Schaaf M. Curriculum development for the workplace using entrustable professional activities (EPAs): AMEE Guide No. 99. *Med Teach.* 2015;37(11):983–1002. doi:10.3109/0142159X.2015.1060308.
14. Warm EJ, Mathis BR, Held JD, Pai S, Tolentino J, Ashbrook L, et al. Entrustment and mapping of observable practice activities for resident assessment. *J Gen Intern Med.* 2014;29(8):1177–1182. doi:10.1007/s11606-014-2801-5.
15. Warm EJ, Held JD, Hellmann M, Kelleher M, Kinnear B, Lee C, et al. Entrusting observable practice activities and milestones over the 36 months of an internal medicine residency. *Acad Med.* 2016;91(10):1398–1405. doi:10.1097/ACM.0000000000001292.
16. The American Board of Pediatrics. Entrustable professional activities for subspecialties. <https://www.abp.org/subspecialty-epas#Critcare>. Accessed June 17, 2019.
17. van der Vleuten CP. The assessment of professional competence: developments, research and practical implications. *Adv Health Sci Educ.* 1996;1(1):41–67. doi:10.1007/BF00596229.
18. Downing SM, Haladyna TM. Validity and its threats. In: Downing SM, Yudkowsky R, eds. *Assessment in Health Professions Education*. New York, NY: Routledge Taylor and Francis Group; 2009.
19. American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational and Psychological Testing. *Standards for Educational and Psychological Testing*. Washington, DC: AERA; 2014.
20. Park YS, Riddle J, Tekian A. Validity evidence of residency competency ratings and the identification of problem residents. *Med Educ.* 2014;48(6):614–622. doi:10.1111/medu.12408.
21. Park YS, Zar FA, Norcini JJ, Tekian A. Competency evaluations in the next accreditation system: contributing to guidelines and implications. *Teach Learn Med.* 2016;28(2):135–145. doi:10.1080/10401334.2016.1146607.
22. Park YS, Hicks PJ, Carraccio C, Margolis M, Schwartz A. Does incorporating a measure of clinical workload improve workplace-based assessment scores? Insights for measurement precision and longitudinal score growth from ten pediatrics residency programs. *Acad Med.* 2018;93(11S Association of American Medical Colleges Learn Serve Lead: Proceedings of the 57th Annual Research in Medical Education Sessions):21–29. doi:10.1097/ACM.0000000000002381.
23. Fraenkel JR, Wallen NE, Hyun HH. *How to Design and Evaluate Research in Education*. Boston, MA: McGraw-Hill Education; 2015.
24. Fereday J, Muir-Cochrane E. Demonstrating rigor using thematic analysis: a hybrid approach of inductive and deductive coding and theme development. *Int J Qual Methods.* 2006;5(1):80–92.
25. Beckman TJ, Cook DA, Mandrekar JN. What is the validity evidence for assessments of clinical teaching? *J Gen Intern Med.* 2005;20(12):1159–1164. doi:10.1111/j.1525-1497.2005.0258.x.
26. ten Cate O, Hart D, Ankel F, Busari J, Englander R, Glasgow N, et al. Entrustment decision making in clinical training. *Acad Med.* 2016;91(2):191–198. doi:10.1097/ACM.0000000000001044.
27. Choo KJ, Arora VM, Barach P, Johnson JK, Farnan JM. How do supervising physicians decide to entrust residents with unsupervised tasks? A qualitative analysis. *J Hosp Med.* 2014;9(3):169–175. doi:10.1002/jhm.2150.
28. Teman NR, Gauger PG, Mullan PB, Tarpley JL, Minter RM. Entrustment of general surgery residents in the operating room: factors contributing to provision of resident autonomy. *J Am Coll Surg.* 2014;219(4):778–787. doi:10.1016/j.jamcollsurg.2014.04.019.
29. Sterkenburg A, Barach P, Kalkman C, Gielen M, ten Cate O. When do supervising physicians decide to entrust residents with unsupervised tasks? *Acad Med.* 2010;85(9):1408–1417. doi:10.1097/ACM.0b013e3181eab0ec.
30. Dijksterhuis MG, Voorhuis M, Teunissen PW, Schuwirth LW, ten Cate OT, Braat DD, et al. Assessment of competence and progressive independence in postgraduate clinical training. *Med Educ.* 2009;43(12):1156–1165. doi:10.1111/j.1365-2923.2009.03509.x.
31. Meert KL, Clark J, Eggly S. Family-centered care in the pediatric intensive care unit. *Pediatr Clin North Am.* 2013;60(3):761–772. doi:10.1016/j.pcl.2013.02.011.

32. Curtis JR, White DB. Practical guidance for evidence-based family conferences. *Chest*. 2008;134(4):835–843. doi:10.1378/chest.08-0235.
33. Levetown M, American Academy of Pediatrics Committee on Bioethics. Communicating with children and families: from everyday interactions to skill in conveying distressing information. *Pediatrics*. 2008;121(5):e1441–e1460. doi:10.1542/peds.2008-0565.
34. Holmboe ES. Faculty and the observation of trainees' clinical skills: problems and opportunities. *Acad Med*. 2004;79(1):16–22.
35. Massie J, Ali JM. Workplace-based assessment: a review of user perceptions and strategies to address the identified shortcomings. *Adv in Health Sci Educ Theory Pract*. 2015;20(2):455–473. doi:10.1007/s10459-015-9614-0.
36. Day SC, Grosso LJ, Norcini JJ, Blank LL, Swanson DB, Horne MH. Residents' perception of evaluation procedures used by their training program. *J Gen Intern Med*. 1990;5(5):421–426. doi:10.1007/bf02599432.
37. Holmboe ES, Edgar L, Hamstra S. *Accreditation Council for Graduate Medical Education. The*

Milestones Guidebook. Washington, DC: ACGME; 2016.



Amanda R. Emke, MD, MHPE, is Assistant Professor, Pediatric Critical Care, Washington University in St Louis; **Yoon Soo Park, PhD**, is Associate Professor and Associate Head, Department of Medical Education, and Director of Research, Office of Educational Affairs, University of Illinois at Chicago College of Medicine; **Sushant Srinivasan, MD**, is Assistant Professor, Department of Pediatrics, University of Wisconsin School of Medicine and Public Health; and **Ara Tekian, PhD, MHPE**, is Professor, Department of Medical Education, University of Illinois at Chicago College of Medicine.

Funding: The authors report no external funding source for this study.

Conflict of interest: The authors declare they have no competing interests.

The results were presented at the Society of Critical Care Medicine 47th Critical Care Congress, San Antonio, Texas, February 25–28, 2018.

Corresponding author: Amanda R. Emke, MD, MHPE, Washington University in St Louis, NWT CB 8116, One Children's Place, St Louis, MO 63110, 314.454.2678, emke_a@wustl.edu

Received January 23, 2019; revisions received April 22, 2019, and May 29, 2019; accepted May 29, 2019.