



## Regular Article

# Leveraging faculty development to support validation of entrustable professional activities assessment tools in anatomic and clinical pathology training

Bronwyn H. Bryant, MD<sup>a,\*</sup>, Scott R. Anderson, MD<sup>a</sup>, Mark Brissette, MD<sup>b</sup>, John M. Childs, MD<sup>c</sup>, Dita Gratzinger, MD, PhD<sup>d</sup>, Kristen Johnson, PhD<sup>e</sup>, Deborah E. Powell, MD<sup>f</sup>, Suzanne Zein-Eldin Powell, MD<sup>g</sup>, Charles F. Timmons, MD, PhD<sup>h</sup>, Cindy B. McCloskey, MD<sup>i</sup>

<sup>a</sup> University of Vermont Medical Center, Department of Pathology and Laboratory Medicine, Burlington, VT, USA

<sup>b</sup> University of Colorado Anschutz Medical Campus, Department of Pathology, Aurora, CO, USA

<sup>c</sup> Geisinger Medical Center, Department of Pathology, Danville, PA, USA

<sup>d</sup> Stanford University School of Medicine, Department of Pathology, Stanford, CA, USA

<sup>e</sup> College of American Pathologists, Northfield, IL, USA

<sup>f</sup> University of Minnesota Medical School, Department of Laboratory Medicine and Pathology, Minneapolis, MN, USA

<sup>g</sup> Houston Methodist Hospital, Department of Pathology and Genomic Medicine, Houston, TX, USA

<sup>h</sup> UT Southwestern Medical Center, Department of Pathology, Dallas, TX, USA

<sup>i</sup> University of Oklahoma College of Medicine, Department of Pathology, Oklahoma City, OK, USA

## A B S T R A C T

Entrustable professional activities (EPAs) are observable activities that define the practice of medicine and provide a framework of evaluation that has been incorporated into US medical school curricula in both undergraduate and graduate medical education. This manuscript describes the development of an entrustment scale and formative and summative evaluations for pathology EPAs, outlines a process for faculty development that was employed in a pilot study implementing two Anatomic Pathology and two Clinical Pathology EPAs in volunteer pathology residency programs, and provides initial validation data for the proposed pathology entrustment scales. Prior to implementation, faculty development was necessary to train faculty on the entrustment scale for each given activity. A “train the trainer” model used performance dimension training and frame of reference training to train key faculty at each institution. The session utilized vignettes to practice determination of entrustment ratings and development of feedback for trainees as to strengths and weaknesses in the performance of these activities. Validity of the entrustment scale is discussed using the Messick framework, based on concepts of content, response process, and internal structure. This model of entrustment scales, formative and summative assessments, and faculty development can be utilized for any pathology EPA and provides a roadmap for programs to design and implement EPA assessments into pathology residency training.

**Keywords:** Assessment, Entrustable professional activities, Faculty development, Pathology, Validation

## Introduction

Entrustable professional activities (EPAs) are becoming increasingly common in the landscape of competency-based medical education (CMBE). EPAs, by their design, reflect the activities physicians perform each day that define their specialty. This framework, anchored in entrustment, reflects the goals of residency training—graduating competent physicians who are entrusted to practice independently in their field.<sup>1</sup> There is a well-established and widely adopted framework for the description of EPAs, and many fields have defined and published EPAs along with their experiences in implementation of the framework.<sup>2–7</sup> For

the specialty of pathology, EPAs were first defined for primary certification,<sup>8,9</sup> followed by dedicated EPAs for fellowship training.<sup>10,11</sup>

The Pathology National EPA Pilot study is a project of the EPA Working Group, a team co-sponsored by the Association of Pathology Chairs (APC) and College of American Pathologists (CAP) and formed to advance EPAs in pathology residency training. As part of the Pathology National EPA Pilot, the goal of this specific work was twofold. The first goal was to develop EPA-based formative and summative assessment tools for four pathology EPAs, including defining entrustment scales. Our driving principles were to provide tools that are specific enough to align with the day-to-day practice of pathology while flexible enough to be

Refers to: <https://doi.org/10.1016/j.acpath.2024.100110>.

\* Corresponding author. University of Vermont Medical Center Department of Pathology and Laboratory Medicine Larner College of Medicine 111 Colchester Avenue Burlington, VT 05401 USA.

E-mail address: [Bronwyn.Bryant@uvmhealth.org](mailto:Bronwyn.Bryant@uvmhealth.org) (B.H. Bryant).

<https://doi.org/10.1016/j.acpath.2024.100111>

Received 25 October 2023; Received in revised form 26 January 2024; Accepted 3 February 2024; Available online xxxx

eISSN2374-2895/© 2024 The Author(s). Published by Elsevier Inc. on behalf of Association of Pathology Chairs. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

adapted to pathology residency programs nationwide. The second goal was to perform validity work for the evaluation tools as part of the faculty development program for a planned pilot study to implement EPA evaluations at multiple pathology residency programs.

Significant variation exists in the implementation of entrustment scales for EPAs across medical specialties and certification organizations. In ten Cate's model, the entrustment scale includes anchors such as "Not allowed to practice EPA," "Allowed to practice EPA unsupervised," or "Allowed to supervise others in practice of EPA."<sup>12</sup> These descriptors clearly designate at what level of autonomy the trainee is able to practice the defined EPA, including their ability to oversee other trainees. Interestingly, some specialties have modified this scale to shift the perspective from the ability of the trainee to perform the activity under various levels of supervision to the behavior characteristics of the faculty required to supervise the trainee. This language may be more intuitive to faculty as it reflects their behavior as a teacher and supervisor. For example, in anesthesiology, Weller et al. described supervision in terms of a faculty member's comfort in leaving the theater during a procedure.<sup>13</sup> A similar example from surgery is that of the Ottawa Surgical Competency Operating Room Evaluation (O-SCORE) Scale, where faculty behaviors range from "I had to do" to "I had to prompt them from time to time" to "I did not need to be there."<sup>14</sup> Aside from medical specialty-specific entrustment scale language, further variation exists in the supervision language of certification organizations such as the Accreditation Council for Graduate Medical Education (ACGME), where supervision is described using the terms "Direct," "Indirect," and "Oversight" supervision. Theoretically, the level of assessed entrustment should define the necessary supervision, and therefore, these concepts should be conceptually aligned.

Given the complexity and overlapping nature of language for assessment and supervision, faculty training and validation of assessment scales are critical and are often overlooked steps of assessment scale implementation both in practice and in the literature. Various models for determining the validity of these tools for their intended use have been described.<sup>15-17</sup> One such model, the Messick framework, uses the concepts of content, response process, internal structure, relationship to other variables, and consequences to support "construct validity." Construct validity is the concept that an evaluation score represents the intended underlying construct, which in the case of EPAs is trainee competence as defined through entrustment.<sup>15</sup> Using aspects of the Messick framework, this paper describes the development of the formative and summative assessment tools and the faculty development sessions used to collect evidence of validity of the previously published entrustment scales that were adapted for pathology residency training.

## Materials and methods

The purpose of this project was to develop formative and summative evaluation tools to be used in a pilot study examining the utility of ten Cate's Entrustable Professional Activities framework in primary anatomic and clinical pathology training, including establishment of entrustment scales for pathology. As part of this effort, we also performed a preliminary validation of the entrustment scales as a component of the faculty development phase of the EPA pilot study. Institutional Review Board review of the EPA pilot study was obtained through the University of Vermont.

### *Entrustment scale and EPA assessment tool development*

Four EPAs were selected from the previously published list of EPAs for primary anatomic pathology/clinical pathology (AP/CP) certification.<sup>8</sup> These EPAs were chosen as they cover a mix of AP and CP topics, and most residency programs were likely to have a sufficient clinical volume in these practice areas for the pilot study.

- (1) Perform a medical autopsy [subdivided into 1) gross procedure through Preliminary Anatomic Diagnoses (PAD) and 2) Final Anatomic Diagnoses (FAD)].

- (2) Perform an intraoperative consultation/frozen section.
- (3) Evaluate and report an adverse event involving transfusion of blood components.
- (4) Compose a diagnostic report for clinical laboratory testing (peripheral blood smear review) requiring pathology interpretation.

To develop pathology-specific entrustment scales, we performed a literature review to identify published entrustment scales that could be adapted for use in pathology-specific activities, adopting language and concepts from the O-SCORE and Ottawa Clinic Assessment Tool (OCAT) scales.<sup>14,18</sup> Published scales were compared to the ten Cate original and expanded entrustment scales and were also evaluated within the context of the 2018 ACGME definitions of supervision (most current version at the time of development).<sup>19</sup> The entrustment scales were used as a component of both the formative and summative assessments.

In addition to the overall entrustment scales, we also included an assessment of the knowledge and skills statements for each EPA as part of the formative assessment process. The goals of including knowledge and skills statements in the formative assessment process were to reinforce the shared mental model of competence for each EPA among faculty and residents, to prompt faculty on items to assess, and to provide faculty with specific components of the activity they could address with formative feedback. Based on concerns about the cognitive load of the evaluation tool, it was felt that a global competency rating for each knowledge and skills statement was most appropriate. Multilevel behavioral anchors for each knowledge and skills statements would add considerably to the amount of reading required for faculty to complete what was designed to be an efficient formative assessment tool (e.g., an EPA with five levels of assessments for five knowledge and skills statements would require 25 behavioral anchor descriptions). Also, the knowledge and skills statements themselves are written to describe competent performance of each function and provide some guidance as to behavioral expectations. For the knowledge and skills statements, a competency-based assessment scale was adapted from Nousiainen et al.<sup>20</sup> This descriptive, 5-point Likert scale was anchored in competency language and provided specific feedback in the formative setting, even in the absence of narrative feedback.

As a final component of the development process, members of the working group shared final drafts of the assessment tools with local content experts (both experts at the EPA task and experts at training residents in said task). Comments on content and wording of knowledge and skill statements were incorporated into the final EPA assessment tools.

The intended use of the entrustment scale is in the implementation of formative EPAs in anatomic and clinical pathology residency programs. A separate manuscript, "National pilot of entrustable professional activities in pathology residency training", describes in detail the pilot implementation of four EPAs together with stakeholder feedback.<sup>23</sup> The structure of the pilot implementation study is described here in brief as the intended use of the entrustment scale as an assessment tool is within the context of EPA implementation. The formative EPAs (including knowledge and skill statements, entrustment scale, and narrative feedback) were designed to be used as a workplace-based assessment, integrated into clinical workflow. Trainees would request 2-3 formative assessments each week on pertinent clinical rotations (or during on-call activities). At the end of a rotation (or in preparation for a semi-annual clinical competency committee review), a designee would review the EPAs completed to date and complete a summative EPA, which included a summative entrustment rating and brief highlight of narrative feedback. This summative EPA was mapped to milestones.<sup>21</sup>

### *Faculty development and entrustment scale validation*

Faculty development utilized a "train-the-trainer" model. One AP and two CP sessions were facilitated by authors B.H.B. (AP) and M.B. (CP). Program directors, associate program directors, and/or rotation directors

were invited to attend AP- and/or CP-specific training sessions held on Zoom (Zoom Video Telecommunications, San Jose, CA), which allowed for use of polling technology built into the platform. Participants were asked to watch a short background video ahead of the session. The live session included a 5-min review of the background video, including defining EPAs, linking EPAs to ACGME milestones/competencies, describing components of trust, and listing the purpose of faculty development. The practical application of the training utilized performance dimension training and frame of reference training. Performance dimension training involved asking faculty to list essential steps to perform each EPA, thereby recreating the specific EPAs in their own words and helping to create a shared mental model of the activity. Frame of reference training involved reviewing two fictional clinical vignettes for each EPA. Briefly, the vignettes used were written by EPA working group members or subject matter experts involved in pathology graduate medical education. EPAs and entrustment scales were provided to writers with instructions to create vignettes targeting entrustment level 2 (“I had to talk them through”) and level 4 (“I needed to be available just in case”). Finally, vignettes were reviewed by select working group members prior to utilization in faculty development activities. During the faculty development activity, faculty reviewed each individual vignettes and were polled to provide a single entrustment rating for the depicted resident. This was followed by a poll about the components of trust that were positively demonstrated and the components of trust that needed improvement. The poll responses were shared with the group, and participants were asked to discuss the reasons behind poll responses. The facilitator also discussed the entrustment level targeted by the vignette author. Poll results were exported from the recorded session and analyzed for trends based on target rating of each vignette. As the number of respondents varied between polls, results are presented as a percent of respondents. The number of responses is listed in the figure legends.

**Results**

**EPA entrustment scales**

The working group considered entrustment scales that would be intuitive, replicable, and useful for four common pathology professional

activities across anatomic and clinical pathology; the two anatomic pathology activities involve procedures and reporting (autopsy and frozen section), whereas the two clinical pathology activities involve clinical management and reporting (transfusion adverse event and peripheral smear/diagnostic report for clinical laboratory test). To align faculty supervision behaviors with the language used in the entrustment scales, the working group developed entrustment scales for the categories of reporting, clinical management, and performance of procedures (Table 1).

**Structure of formative and summative assessment tools**

As described above, formative and summative evaluations were developed for each EPA to be used in the Pathology National EPA Pilot study. Components of the formative assessment included a rating of case complexity, other case characteristics as needed, performance ratings of knowledge and skills statements, solicitation of free-text feedback asking for specific examples of aspects of the EPA that were performed well and aspects of the EPA that could be improved upon, and a final entrustment score for the EPA based on the observed encounter. Fig. 1 demonstrates the formative evaluation of the performance of an intraoperative consultation or frozen section. All formative evaluations can be found as Supplemental Materials 1. The summative assessment tool was simpler than the formative tool and included a summative entrustment determination and summative free-text feedback. The summative tool also included a key mapping to 2018 ACGME supervision defined in the ACGME Program Requirements<sup>19</sup> and to specific ACGME Milestones ratings<sup>21</sup> at each entrustment level for each EPA. Fig. 2 shows the summative assessment of the performance of an intraoperative consultation or frozen section. All summative evaluations can be found as Supplemental Materials 2.

**Faculty development validity study**

Twenty pathologists attended AP faculty development training, and 21 pathologists attended CP faculty development training (Table 2). Faculty responses were collected in reaction to vignettes developed for AP (two EPAs/four vignettes total) and/or CP (two EPAs/four vignettes total). Vignettes are provided as Supplemental Materials 3. For each

**Table 1**  
Comparison of supervision and entrustment scales.

ACGME supervision levels (ACGME Common Program Requirements Effective July 1, 2018)	Entrustment scales adapted for pathology from O-SCORE <sup>14</sup> and OCAT <sup>18</sup>		
	Procedure-based EPAs (i.e., autopsy procedure, frozen section)	Situation based/clinical problem-solving (i.e., managing transfusion reaction)	Reporting based (i.e., peripheral smear report, autopsy reporting)
Observation only—not allowed to practice EPA (not recognized by ACGME)	“I had to do”—requires complete guidance, unprepared to do, or had to do for them	“I had to manage”—requires complete guidance, unprepared to do, or had to do for them	“I had to draft report”—requires complete guidance, unprepared to do, or had to do for them
Direct supervision—proactive, attending in room	“I had to talk them through”—demonstrates some skill but requires significant direction/oversight	“I had to guide them through”—demonstrates some problem solving/reporting ability but requires significant direction, oversight, and/or report editing	“I had to co-write or significantly edit/correct the report”—able to draft a rudimentary report only
Indirect supervision—reactive/on-demand, attending available in person or remotely (likely in person)	“I had to direct them from time to time”—demonstrates some independence but requires intermittent prompting and/or double checking of work	“I had to prompt them from time to time”—demonstrates some independence with problem solving/reporting, but requires intermittent prompting, double checking of work, and/or report editing	“I had to offer suggestions and/or edit the report”—able to draft majority of report, but requires some editing or correction of work
Indirect supervision—reactive/on-demand, attending available in person or remotely (likely remotely)	“I needed to be available just in case”—independence but may need assistance with nuances of certain cases or skills; still requires supervision for safe practice (may be remote)	“I needed to be available just in case”—independence but may need assistance with nuances of certain situations or reporting; still requires supervision for safe practice (may be remote)	“I needed to review the report just in case”—able to draft complete report that requires minimal editing
Oversight supervision—post hoc feedback, review after the fact	“I did not need to be there”—Complete independence, understands risks and performs safely, practice ready	“I did not need to be there”—Complete independence, understands risks and performs safely, practice ready	“I did not need to review report”—able to draft complete and polished report that can be signed out as is

ACGME: Accreditation Council for Graduate Medical Education; EPA: entrustable professional activity; OCAT: Ottawa Clinic Assessment Tool; O-SCORE: Ottawa Surgical Competency Operating Room Evaluation.

## Formative assessment: Intraoperative Consult/Frozen Section

### Case Complexity Rating:

How would you describe the complexity of the case evaluated?

- Low
- Medium
- High

### EPA Formative Assessment:

Using the scale descriptors below, please rate the level of performance for the following Knowledge and Skills Statements for the case observed. If a skill was not applicable, not observed, or performed by someone other than the resident (i.e. technologist, PA, etc.) choose N/A.

### Scale Descriptors

Fails to Meet Essential Competencies 1	2	Meets Essential Competencies 3	4	Demonstrates Enhanced Competencies 5	U/A
<ul style="list-style-type: none"> <li>• Below the minimally acceptable level for competence</li> <li>• Did not know limits and/or ask for assistance when needed</li> <li>• Unable to manage work in timely manner</li> <li>• Didn't improve with or respond to feedback</li> </ul>		<ul style="list-style-type: none"> <li>• Demonstrated a solid understanding of the issues, interpretation of problems and basic implementation of solution(s)</li> <li>• Knew and worked within competence limits</li> <li>• Did what was expected in a timely way</li> <li>• Responded to or improved with feedback</li> </ul>		<ul style="list-style-type: none"> <li>• Demonstrated excellence in his or her understanding of the issues, ability to interpret problems and implement solutions</li> <li>• Exceeds the benchmark for competent performance</li> <li>• Demonstrated ease and efficiency in handling work</li> <li>• Anticipated areas needing refinement and strove for continuous improvement</li> </ul>	

Fig. 1. Formative assessment for performance of intraoperative consultation/frozen section.

vignette, data collected included an entrustment rating and “positive” and “needs improvement” reactions to components of trust. Not all participants answered every poll question; therefore, data are graphed as the percent of responses.

*AP (procedural) topics:* For the autopsy vignettes, entrustment ratings clustered around the bottom three entrustment ratings for the level 2 vignette (Fig. 3A) and the top two entrustment ratings for the level 4 vignette (Fig. 3B). For the frozen section vignettes, entrustment ratings

were clustered in the bottom two entrustment ratings for the level 2 vignette (Fig. 3C) and the top three entrustment ratings for the level 4 vignette (Fig. 3D).

*CP (reporting and clinical management) topics:* For the peripheral smear review vignettes, entrustment ratings were clustered in the bottom two entrustment ratings for the level 2 vignette (Fig. 4A) and the top three entrustment ratings for the level 4 vignette (Fig. 4B). For the transfusion reaction vignette, entrustment ratings were clustered in the bottom three

**Performance Rating:**

Knowledge/Skills Statement	1	2	3	4	5	N/A
<b>1. Properly identify patient specimen before and throughout the frozen section/intraoperative consultation process</b>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>2. Determine correct method of tissue preparation needed (e.g., frozen section, cytologic preparation, gross evaluation, etc.)</b>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>3. Sample tissue appropriately for diagnosis and/or evaluation of margins</b>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>4. Prepare (e.g., cut frozen sections on cryostat, perform cytologic touch preparations) and stain slides for microscopic evaluation</b>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>5. Evaluate and interpret microscopic findings</b>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>6. Communicate diagnosis, including limitations of technique or interpretation in a timely manner</b>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>7. Apportion tissue for special studies if needed and prepare tissue in appropriate fixative or media for transport, storage, or further processing</b>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**Free Text Formative Feedback:**

1. Give at least 1 specific aspect of the EPA performed well based on review of the Formative Assessments.
2. Give at least 1 specific suggestion for improvement based on review of the Formative Assessments.

Fig. 1. (continued).

**Entrustment Determination:**

Based on your current observations of and experience to-date with this resident, please rate the **level of supervision** you suggest for this resident to perform an intraoperative consultation/frozen section:

<b>ACGME Supervision Levels</b>	<b>Observation Only</b> -not allowed to practice EPA	<b>Direct Supervision</b> -proactive -attending in the room	<b>Indirect Supervision</b> -reactive/on-demand -attending available in person or remotely (likely in person)	<b>Indirect Supervision</b> -reactive/on-demand -attending available in person or remotely (likely remotely)	<b>Oversight Supervision</b> -post hoc feedback -review after the fact
<b>Entrustment Scale</b>	<b>"I had to do"</b> - Requires complete guidance, unprepared to do, or had to do for them	<b>"I had to talk them through"</b> - Demonstrates some skill but requires significant direction/over sight	<b>"I had to direct them from time to time"</b> - Demonstrates some independence, but requires intermittent prompting and/or double checking of work	<b>"I needed to be available just in case"</b> - Independence but may need assistance with nuances of certain cases or skills; still requires supervision for safe practice (may be remote)	<b>"I did not need to be there*"</b> - Complete independence, understands risks and performs safely, practice ready (*ACGME oversight supervision required for all trainees)
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Fig. 1. (continued).

entrustment ratings for the level 2 vignette (Fig. 4C) and the spanned all entrustment ratings for the level 4 vignette, with 47% picking level 4 (Fig. 4D).

*Components of trust:* For level 2 vignettes, the most commonly cited concerns about the trainee's performance were a lack of competence and/or reliability (Fig. 5A). Conversely, for level 4 vignettes, trainees were commonly rated as performing well in regard to competency and/or reliability (Fig. 5B). If any component of trust was performed well by the trainees in level 2 vignettes, it was most often humility and honesty, a trainee being honest about what they did not know and humble enough to admit it; although, several respondents noted "none" of the components of trust were performed well. When asked if level 4 vignette trainees had any areas to work on, "N/A, I trust this resident" was most commonly selected.

**Discussion**

EPAs have been particularly well documented and validated for procedure-based tasks (especially in surgical specialties).<sup>14,22</sup> In pathology, the entrusted skill can vary quite broadly from a procedure, to writing a report, to managing a clinical event or activity. This study demonstrated a framework for adjusting entrustment scales to match procedure-based tasks, clinical situation/management tasks, and written report tasks. Through the faculty development training for the Pathology National EPA Pilot study, we were able to provide evidence supporting the construct validity of the entrustment scale for each of the four EPAs.

The alignment of responses demonstrated reasonable understanding by participants as to the meaning of the scale.

Following the Messick framework for validity and reliability in psychometric instruments,<sup>15</sup> such as entrustment scale, this study showed strengths in the areas of *content*, *response process*, and *internal structure*. The other two factors, *relationship to other variables* and *consequences*, are not applicable at this time as participants were assessing fictional vignettes; however, future studies are planned to compare actual EPA data to ACGME Milestones and to model EPA data over time to better study consequences of the assessment.

In terms of content, EPAs are well aligned with the intended construct of the assessment tool. Knowledge and skill statements define the intended use, or what the rater (faculty) should be looking for in the resident's performance. Other supports of the *content* domain include the preliminary review by experts in each field, who know not just the details of performing the task but are also well versed in supervising and training residents at each task.

The faculty development model supports the domain of *response process*. The performance dimension training section showed that faculty generated a similar list of knowledge and skill statements as those generated by the working group, showing that all participants have a reasonably similar idea of what competent performance should look like. Providing faculty an opportunity to discuss ratings with each other during training can help align interpretations of observations. Periodic "calibration" of faculty with more training scenarios (performed during the subsequent Pathology National EPA Pilot study and future EPA

## Summative Assessment: Intraoperative Consult/Frozen Section

### Summative Entrustment Determination:

The following question is to be completed at the end of the rotation by the designated faculty member, rotation director, or program director based on a review of all Formative Assessments completed during **this rotation**. ONLY THIS QUESTION will map to Milestones (see below)

ACGME Supervision Levels	Observation Only -not allowed to practice EPA	Direct Supervision -proactive -attending in the room	Indirect Supervision -reactive/on-demand -attending available in person or remotely (likely in person)	Indirect Supervision -reactive/on-demand -attending available in person or remotely (likely remotely)	Oversight Supervision -post hoc feedback -review after the fact
Entrustment Scale	"I had to do" - Requires complete guidance, unprepared to do, or had to do for them	"I had to talk them through" - Demonstrates some skill but requires significant direction/oversight	"I had to direct them from time to time" - Demonstrates some independence, but requires intermittent prompting and/or double checking of work	"I needed to be available just in case" - Independence but may need assistance with nuances of certain cases or skills; still requires supervision for safe practice (may be remote)	"I did not need to be there*" - Complete independence, understands risks and performs safely, practice ready (*ACGME oversight supervision required for all trainees)
Maps to Milestone:	PC1 = NYCL1 PC3 = NYCL1 PC4 = NYCL1 PC5 = 1.0 MK2 = NYCL1	PC1 = 1.0 PC3 = 1.0 PC4 = 1.0 PC5 = 2.0 MK2 = 1.0	PC1 = 2.0 PC3 = 2.0 PC4 = 2.0 PC5 = 3.0 MK2 = 2.0	PC1 = 3.0 PC3 = 3.0 PC4 = 3.0 PC5 = 3.5 MK2 = 3.0	PC1 = 4.0 PC3 = 4.0 PC4 = 4.0 PC5 = 4.0 MK2 = 4.0

NYCL1 – Not Yet Completed Level 1

PC1: Reporting

PC3: Clinical Consultation, including On-Call Interactions

PC4: Interpretation and Diagnosis

PC5: Intra-Operative Consultation, including Frozen Section

MK2: Clinical Reasoning

### Summative Evaluation Free Text Comments:

1. Give at least 1 specific aspect of the EPA performed well based on review of the Formative Assessments.
2. Give at least 1 specific suggestion for improvement based on review of the Formative Assessments.

Fig. 2. Summative assessment for performance of intraoperative consultation/frozen section.

studies) will provide further evidence in support of good *response process*.

For the domain of *internal structure*, we report an acceptable consistency and reliability to the faculty ratings, even though faculty did not all agree on a single entrustment level for each vignette. Faculty gave a range of ratings to include one level above and one level below each

target entrustment rating. Considering faculty were working with fictional vignettes without any prior “experience” with the trainee, some variance is expected. In addition, there is known variability in the relative leniency or severity of faculty (hawks vs. doves). When comparing the two vignettes side by side, there was no overlap in ratings for three of the four EPAs. The variance was higher than expected for one transfusion

**Table 2**  
Faculty development session.

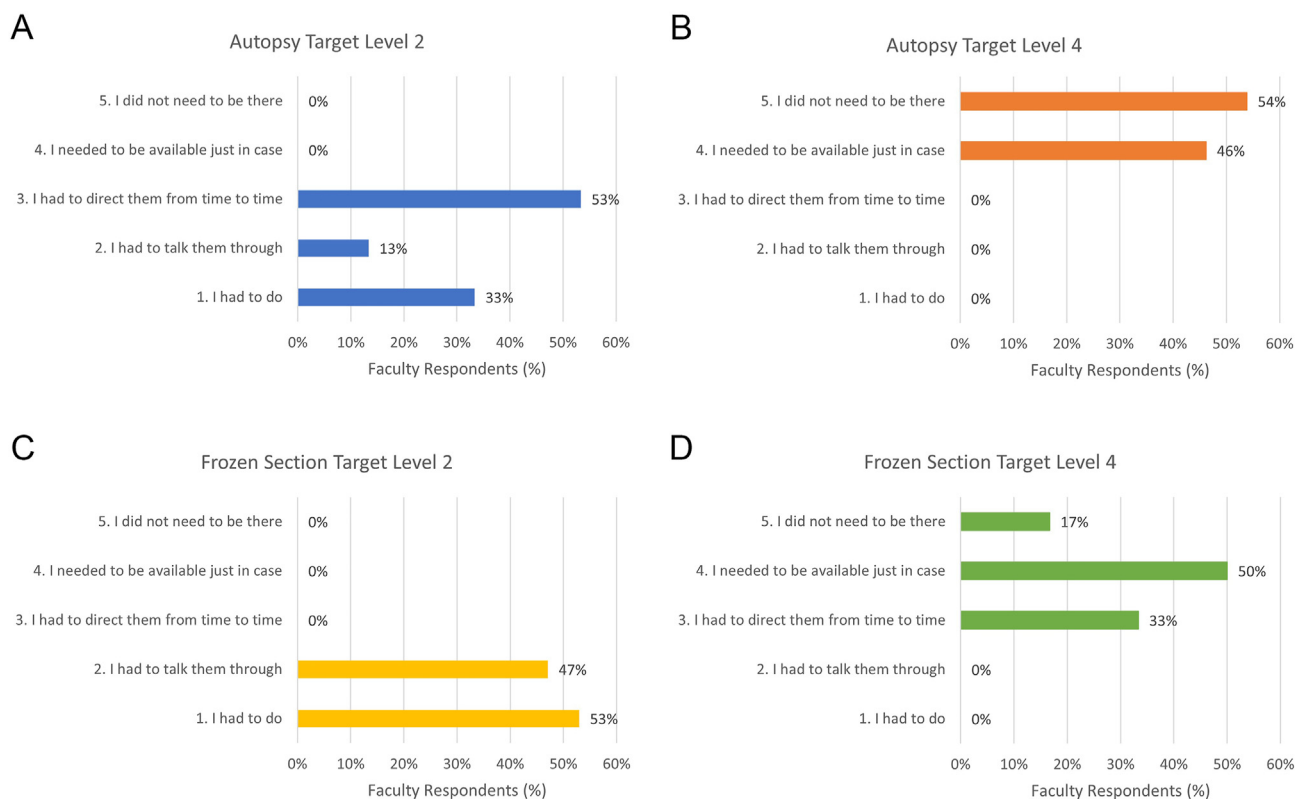
Pre-recording	10-min background video covering the following: <ul style="list-style-type: none"> <li>• Definition of EPAs</li> <li>• Components of trust</li> <li>• Workplace-based assessment</li> <li>• Factors leading to low-quality assessments</li> <li>• Importance of faculty development</li> </ul>
Live session	<p>Pre-assignment review:</p> <ul style="list-style-type: none"> <li>• Definition of EPAs</li> <li>• Components of trust</li> </ul> <p>For each EPA:</p> <ul style="list-style-type: none"> <li>• Performance dimension training                     <ul style="list-style-type: none"> <li>o As a group, define knowledge, skills, and attitudes necessary for the specified activity</li> <li>o Share EPA description for the specified activity and compare with group consensus</li> </ul> </li> <li>• Frame of reference training                     <ul style="list-style-type: none"> <li>o Read level 2 vignette, assign entrustment score, and evaluate components of trust                             <ul style="list-style-type: none"> <li>• Discuss faculty assessment including entrustment score, components of trust, and positive and constructive feedback for the trainee</li> <li>• Develop a consensus on behaviors of a trainee at a lower entrustment level</li> </ul> </li> <li>o Read level 4 vignette, assign entrustment score, and evaluate components of trust                             <ul style="list-style-type: none"> <li>• Discuss faculty assessment including entrustment score, components of trust, and positive and constructive feedback</li> <li>• Develop a consensus on behaviors of a trainee at a higher entrustment level</li> </ul> </li> <li>o Conclude discussion of specific EPA comparing performance characteristics and resident traits at lower vs. higher levels of entrustment</li> </ul> </li> </ul>

EPA: entrustable professional activity.

medicine vignette, which is primarily attributed to the difficulty of the clinical content in the vignette and lack of subspecialty expertise in transfusion medicine of most participants. During the discussion following this vignette (transfusion reaction vignette #1), it was clear most participants did not fully understand the nuances of the vignette, and there was consensus that the trainee performed the workup and management well. The vignette was revised for a second train-the-trainer session a year later, and agreement was much narrower (data not presented).

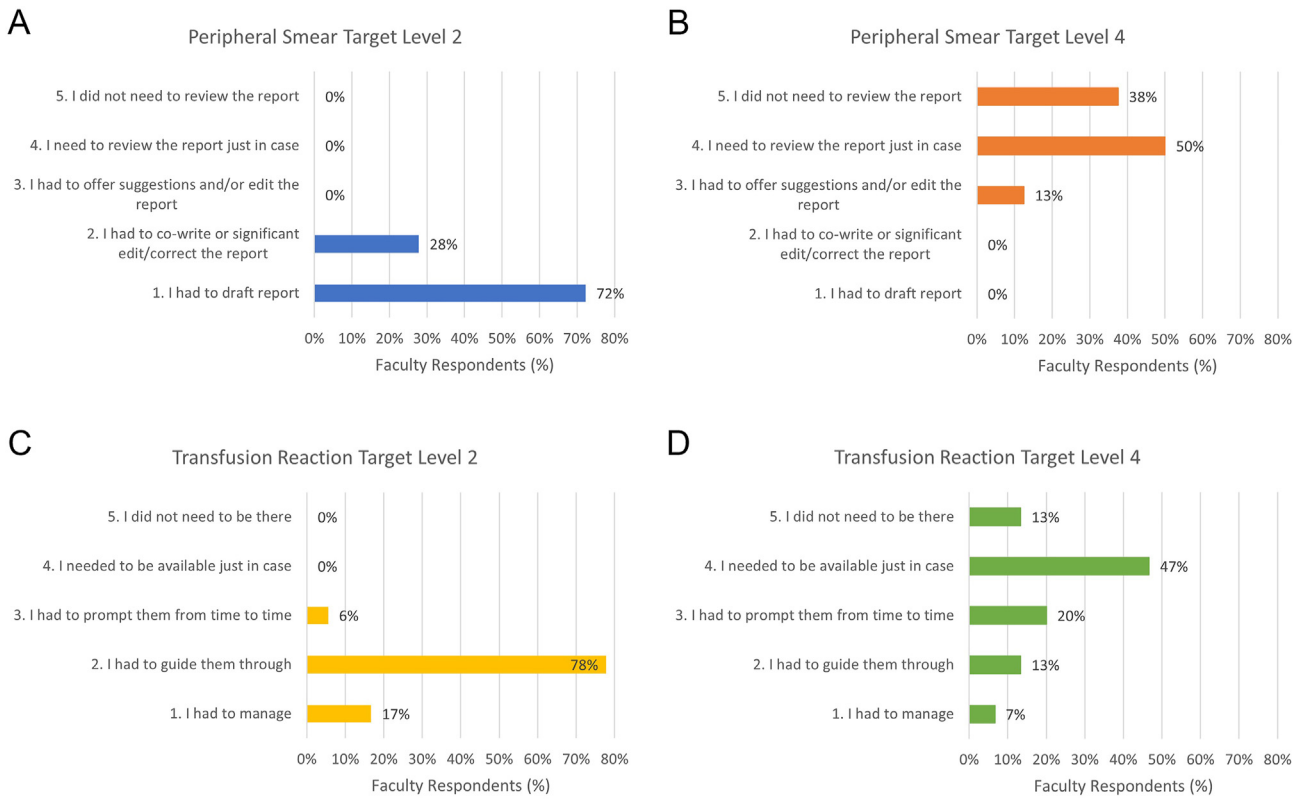
An added benefit to this approach to faculty development is the opportunity to give faculty the language and framework to understand why they selected specific entrustment levels and move beyond gestalt in assigning entrustment. Faculty had the opportunity to discuss both performance of the knowledge and skills needed to perform the EPA and their impressions of performance as it related to the components of trust. By exploring these concepts, the training approach allowed faculty to understand the reasoning behind their entrustment decisions and gave them the language to articulate this decision. This in turn helped faculty understand the potential feedback opportunities when coaching trainees on their performance of an EPA. Cognizance of the decision-making process may also help ward against, or at least recognize, bias in evaluation. For example, in one vignette, a fictional trainee who was underperforming asked for help. One faculty interpreted this in a more trusting light as it demonstrated humility; another faculty interpreted this in a more doubting light, suggesting it showed lack of knowledge. This training provided time and space for faculty to decide on an entrustment level, reflect on what brought them to that decision, and use those insights to provide one piece of effective feedback to the trainee.

One limitation to this study is the inability to perform statistical analysis of agreement for each vignette. Faculty joined from multiple different institutions in the middle of the workday, with multiple faculty

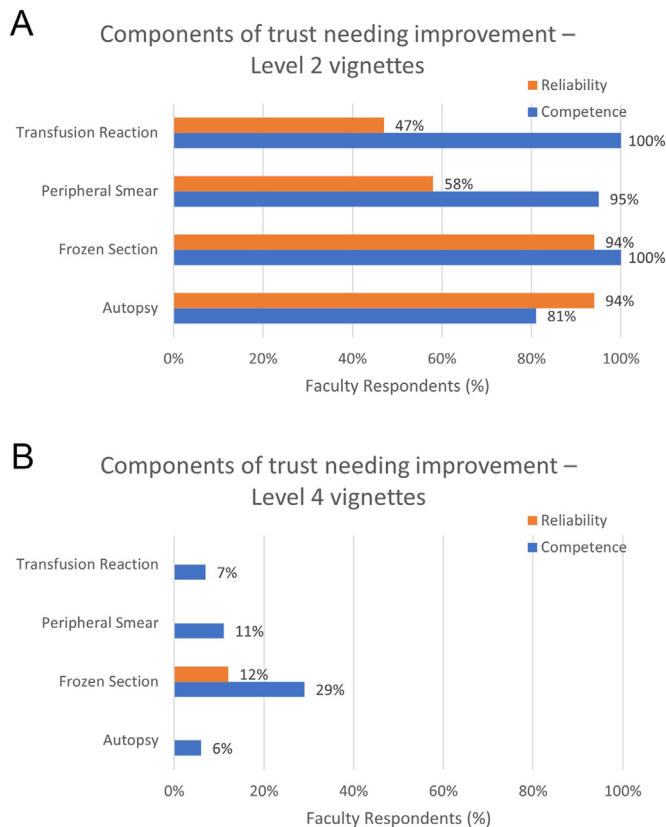


**Fig. 3.** Entrustment ratings for vignettes: (A) autopsy target level 2 (n = 15), (B) autopsy target level 4 (n = 13), (C) frozen section target level 2 (n = 17), and (D) frozen section target level 4 (n = 18).





**Fig. 4.** Entrustment ratings for vignettes: (A) peripheral smear target level 2 (n = 18), (B) peripheral smear target level 4 (n = 16), (C) transfusion reaction target level 2 (n = 18), and (D) transfusion reaction target level 4 (n = 15).



**Fig. 5.** Components of trust (competence and reliability) for improvement by activity for target level 2 (A) and target level 4 (B) [autopsy (n = 16); frozen section (n = 17); peripheral smear (n = 19); transfusion reaction target level 2 (n = 17); transfusion reaction target level 4 (n = 15)].

members attending only a portion of the Zoom call during the discussion, hence the variance in the number of participants for each question. As more programs and faculty participate in EPAs, repeated trainings will generate sufficient data points for more rigorous agreement statistics. Other limitations, as mentioned, include variability in experience of faculty at each session, the fictional nature of the vignettes, and variations in workflow from institution to institution.

In conclusion, we demonstrated how entrustment scale language can be varied slightly to reflect the actions of a specific task in pathology training. Furthermore, we demonstrate a method to provide faculty development in EPAs while at the same time collecting data to support the validity of the entrustment scale. As EPAs are incorporated in pathology and across medical subspecialties, ensuring validity of the assessment tool is a necessary foundation to future research in competency-based medical education.

**Declarations of competing interest**

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Bronwyn H Bryant reports administrative support was provided by the Association of Pathology Chairs. Bronwyn H Bryant reports statistical analysis was provided by the College of American Pathologists. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Funding**

This research is supported by the Association of Pathology Chairs through staff time and the cost of publication, and by the College of American Pathologists through staff time. No internal or external grant funding was used in this study.

## Acknowledgements

The authors would like to acknowledge Cheryl Hanau, MD, Kate Hatlak, EdD, Karen Kaul, MD, PhD, Priscilla Markwood, CAE, Madeline Markwood, Douglas Miller, MD, PhD, Gary Procop, MD, and Cindy Riyad, PhD for their support of this work. Dr. Bryant gratefully acknowledges the guidance of The University of Vermont Larner College of Medicine Teaching Academy.

## Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.acpath.2024.100111>.

## References

1. Ten Cate O, Carraccio C, Damodaran A, et al. Entrustment decision making: extending Miller's pyramid. *Acad Med J Assoc Am Med Coll*. 2021;96(2):199–204. doi:10.1097/ACM.0000000000003800.
2. Ten Cate O, Taylor DR. The recommended description of an entrustable professional activity: AMEE Guide No. 140. *Med Teach*. 2021;43(10):1106–1114. doi:10.1080/0142159X.2020.1838465.
3. Bello RJ, Major MR, Cooney DS, Rosson GD, Lifchez SD, Cooney CM. Empirical validation of the Operative Entrustability Assessment using resident performance in autologous breast reconstruction and hand surgery. *Am J Surg*. 2017;213(2):227–232. doi:10.1016/j.amjsurg.2016.09.054.
4. Aylward M, Nixon J, Gladding S. An entrustable professional activity (EPA) for handoffs as a model for EPA assessment development. *Acad Med*. 2014;89(10):1335–1340. doi:10.1097/ACM.0000000000000317.
5. Hart D, Franzen D, Beeson M, et al. Integration of entrustable professional activities with the milestones for emergency medicine residents. *West J Emerg Med*. 2018;20(1):35–42. doi:10.5811/westjem.2018.11.38912.
6. Hauer KE, Soni K, Cornett P, et al. Developing entrustable professional activities as the basis for assessment of competence in an internal medicine residency: a feasibility study. *J Gen Intern Med*. 2013;28(8):1110–1114. doi:10.1007/s11606-013-2372-x.
7. Stahl CC, Collins E, Jung SA, et al. Implementation of entrustable professional activities into a general surgery residency. *J Surg Educ*. 2020;77(4):739–748. doi:10.1016/j.jsurg.2020.01.012.
8. McCloskey CB, Domen RE, Conran RM, et al. Entrustable professional activities for pathology: recommendations from the College of American pathologists graduate medical education committee. *Acad Pathol*. 2017;4, 237428951771428. doi:10.1177/2374289517714283.
9. Powell DE, Wallschlaeger A. Making sense of the milestones: entrustable professional activities for pathology. *Hum Pathol*. 2017;62:8–12. doi:10.1016/j.humpath.2016.12.027.
10. White K, Qualtieri J, Courville EL, et al. Entrustable professional activities in hematopathology pathology fellowship training: consensus design and proposal. *Acad Pathol*. 2021;8, 237428952199082. doi:10.1177/2374289521990823.
11. Pagano MB, Treml A, Stephens LD, et al. Entrustable professional activities for apheresis medicine education. *Transfusion*. 2020;60(10):2432–2440. doi:10.1111/trf.15983.
12. Ten Cate O. A primer on entrustable professional activities. *Korean J Med Educ*. 2018;30(1):1–10. doi:10.3946/kjme.2018.76.
13. Weller JM, Castanelli DJ, Chen Y, Jolly B. Making robust assessments of specialist trainees' workplace performance. *Br J Anaesth*. 2017;118(2):207–214. doi:10.1093/bja/aew412.
14. Gofton WT, Dudek NL, Wood TJ, Balaa F, Hamstra SJ. The Ottawa Surgical competency operating Room evaluation (O-SCORE): a tool to assess surgical competence. *Acad Med*. 2012;87(10):1401–1407. doi:10.1097/ACM.0b013e3182677805.
15. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. *Am J Med*. 2006;119(2):166. doi:10.1016/j.amjmed.2005.10.036. e7-166.e16.
16. Cook DA, Brydges R, Ginsburg S, Hatala R. A contemporary approach to validity arguments: a practical guide to Kane's framework. *Med Educ*. 2015;49(6):560–575. doi:10.1111/medu.12678.
17. Clauser BE, Margolis MJ, Swanson DB. Issues of validity and reliability for assessment in medical education. In: Holmboe ES, Durning SJ, Hawkins RE, eds. *Practical Guide to the Evaluation of Clinical Competence*. second ed. Elsevier; 2018: 22–36.
18. Rekman J, Hamstra SJ, Dudek N, Wood T, Seabrook C, Gofton W. A new instrument for assessing resident competence in surgical clinic: the Ottawa clinic assessment tool. *J Surg Educ*. 2016;73(4):575–582. doi:10.1016/j.jsurg.2016.02.003.
19. Accreditation Council for Graduate Medical Education. ACGME program Requirements for graduate medical education in anatomic pathology and clinical pathology. Current version of guidelines available at: Accessed October 1, 2019 [https://www.acgme.org/globalassets/pfassets/programrequirements/300\\_pathology\\_2023.pdf](https://www.acgme.org/globalassets/pfassets/programrequirements/300_pathology_2023.pdf); 2018.
20. Nousiainen MT, Mironova P, Hynes M, et al. Eight-year outcomes of a competency-based residency training program in orthopedic surgery. *Med Teach*. 2018;40(10):1042–1054. doi:10.1080/0142159X.2017.1421751.
21. Accreditation Council for Graduate Medical Education. Pathology Milestones, second revision. Updated February. Accessed October 23, 2023 <https://www.acgme.org/globalassets/pdfs/milestones/pathologymilestones.pdf>; 2019.
22. George BC, Teitelbaum EN, Meyerson SL, et al. Reliability, validity, and feasibility of the Zwisch scale for the assessment of intraoperative performance. *J Surg Educ*. 2014;71(6):e90–e96. doi:10.1016/j.jsurg.2014.06.018.
23. Bryant BH, Anderson SR, Brissette M, et al. National pilot of entrustable professional activities in pathology residency training. *Acad Pathol*. 2024. doi:10.1016/j.acpath.2024.100110.