

CHAPTER 11

Monitoring and evaluating EPAs and EPA frameworks

David R. Taylor, Alan Schwartz, Abigail T. Elmes-Patel,
Adrian P. Marty, Marije P. Hennis

Abstract

Evaluation is essential throughout the development of entrustable professional activities (EPAs) to ensure that the resource-intensive work required in EPA development leads to a set (or framework) of EPAs that are fit for their intended purpose. This chapter draws on program evaluation literature to show how EPA core team leaders can conceptualize evaluation as part of the EPA development process. It then covers key evaluation strategies relevant to different stages in EPA development, piloting, and implementation. In addition, an in-depth discussion of the purpose and use of the EQual rubric for evaluating individual EPA quality is provided.

Authors

- David R. Taylor, MD, MHPE. Queen's University, Kingston, Ontario, Canada. Correspondence: taylord@queensu.ca
- Alan Schwartz, PhD, JD. University of Illinois at Chicago, Chicago, USA.

How to cite this book chapter:

Taylor DR, Schwartz A, Elmes AT, Marty AP, Hennis MP. Monitoring and evaluating EPAs and EPA frameworks. In: ten Cate O, Burch VC, Chen HC, Chou FC, Hennis MP. (Eds). *Entrustable Professional Activities and Entrustment Decision-Making in Health Professions Education*, Chapter 11, pp. 121–130. [2024] London: Ubiquity Press. DOI: <https://doi.org/10.5334/bdc.k>

This chapter uses cross-references to other chapters of the same book. For those who read this chapter as a standalone publication: all cross-references can be found at: <https://doi.org/10.5334/bdc>

- Abigail T. Elmes-Patel, PharmD, MHPE. University of Illinois at Chicago, Department of Pharmacy Practice, Chicago, Illinois, USA.
- Adrian P. Marty, MD, MME. University Hospital Balgrist, University of Zurich, Zurich, Switzerland.
- Marije P. Hennis, MD, PhD, MSc Clin Ed, University Medical Center Utrecht, Utrecht, the Netherlands.

Introduction

This chapter describes evaluation methods that can be used during entrustable professional activity (EPA) development initiatives to improve the quality and construction of the EPAs produced. (This is separate from evaluation of learners *performing* EPAs, which is covered in Chapters 17 to 20.) The ultimate goal in EPA development is to produce EPAs that (a) are a clear and representative description of the work of the profession or discipline, (b) are constructed in a way that facilitates teaching and assessment of learners, and (c) can be used in making decisions on readiness for unsupervised practice.¹ Achieving these aims requires an iterative approach to EPA development that adapts to information and opinions collected throughout the process. It follows that an effective evaluation plan for EPA development should go beyond simple adjudication of the quality of proposed EPAs but must also inform the process through which the EPAs are derived, consider the implications (or consequences) of adopting the collective set (or framework) of EPAs, and address ongoing needs for monitoring and revision after implementation.¹ Evaluation in this context therefore includes both formative evaluation approaches (evaluation intended to direct improvements) and summative evaluation approaches (evaluation intended to provide quality assurance for the end product).

The term *EPA framework* is frequently referenced during this chapter and warrants defining from the outset. As described in Chapter 8, a framework of EPAs is a coherent and reasonably comprehensive set of EPAs that define the core activities of a profession or discipline.^{1,2} It should capture all of the essential, entrustable work within the profession's scope of practice. The specific use of the term framework reflects that the collective set of EPAs will be applied differently across different contexts but for a shared purpose and outcome. Consideration of EPA frameworks is important from an evaluation perspective as the cohesiveness and completeness of the collective set of EPAs is an important goal in development.

Conceptualizing evaluation for EPA development

Preparing an effective evaluation plan begins with answering three key questions³: (a) What is the purpose of the evaluation plan? (b) Who is the audience for the evaluation results? (c) What are the questions the evaluation aims to answer? These broad questions have specific considerations when applied to evaluation in EPA development (Table 11.1). It follows from these considerations that the emphasis and purpose of evaluation activities will evolve over the process. Early

Table 11.1: Considerations when creating an evaluation plan for EPAs.

Evaluation questions	Considerations in EPA development
<i>What is the purpose of the evaluation plan?</i>	<ul style="list-style-type: none"> • Direct improvements during development (formative evaluation) • Judge the quality of the product in development and the final end product (formative and summative evaluation) • Support policy decision-making (summative evaluation)
<i>Who is the audience for the evaluation results?</i>	<ul style="list-style-type: none"> • EPA core development team • End users (learners, clinical teachers and educators) • Regulators and accreditors • Public
<i>What are the questions the evaluation aims to answer?</i>	<ul style="list-style-type: none"> • Are the EPAs representative of the core work practiced in the profession? • Does the construction of the EPAs meet quality standards? • Are the EPAs suitable for adoption across programs? • Are the EPAs acceptable from a policy standpoint?

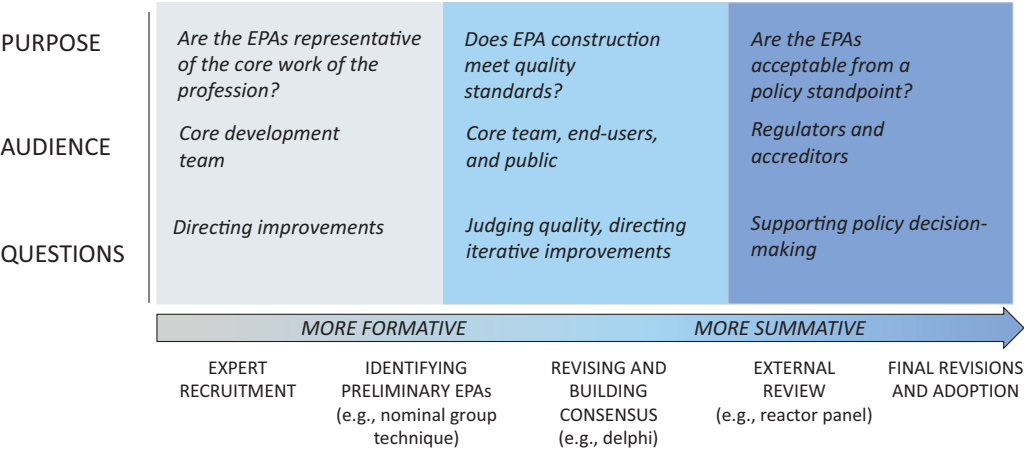


Figure 11.1: Mapping evaluation plan onto EPA development plan.

evaluation activities will be predominantly formative in nature—they aim to improve the process and support product evolution and improvement. As one nears the end of EPA development, the emphasis shifts to quality control of the end product and its suitability for use in high-stakes decisions by regulatory and certifying bodies.

Planning for evaluation starts with mapping the key considerations highlighted in Table 11.1 onto the planned EPA development methods (see Figure 11.1). At first glance, this can make evaluation appear overly complicated and intimidating. But, by taking a step back, it can be appreciated how this mapping exercise clarifies the goals of the work at each phase of the project and provides assurance of readiness for the next phase. For each phase of the project, it lays out:

- What information needs to be gathered?
- Who needs this information?
- How will they use it?

The scope of the evaluation plan should be proportional to the scope of the EPA initiative and aligned to available resources. Smaller local initiatives may choose to focus primarily on evaluating the quality of EPA construction. Initiatives at a national level with implications for licensing decisions will likely require a more comprehensive plan, potentially including an external evaluator. Each section of this chapter describes evaluation approaches useful to EPA development; it is up to project leaders to select and prioritize the most relevant to their initiative.

Evaluation strategies early in development

As the EPA development process begins, the focus of evaluation is formative. The two primary interests are ensuring the clinical content areas covered by the EPAs are representative of the profession and that EPA construction aligns with quality standards for EPAs.

Ensuring the clinical content areas of the EPAs are representative is an evaluation judgment made without prespecified metrics or tools. It aims to ensure that the final collective set (or framework) of EPAs is a cohesive and complete description of the core work done in the profession. At the early stages of EPA development, the focus is simply on establishing that the development team membership is representative of the profession or discipline for which the EPAs are being developed. There is never a perfect composition for any EPA development team, but a diversity of

experts that in broad strokes represents the diversity of practice is important. This is a key checkpoint prior to starting the actual EPA development work. As the development process proceeds and candidate EPAs are proposed and revised, the evaluation questions related to content will become more specific—are there core tasks in the profession (i.e., EPAs) that have been missed? Does the collective set of EPAs over- or underrepresent certain aspects of practice? These questions are often answered through open discussion within the core development team, although they can also be worked into consensus methods such as the Delphi technique. It is important to note that content evaluation is easily overlooked at early stages of EPA development; explicitly including this in the evaluation workplan is important. Content problems in the EPA framework that are only discovered at the end of the project can be difficult to correct.

The other main evaluation interest early in development is the quality of construction for EPAs. Developing EPAs is resource-intensive, so it is important to address quality problems early in the process (see Chapter 9). If EPAs with construction problems move into and through consensus methods, the construction issues can negatively impact consensus ratings, leading to rejection of proposed EPAs that, in fact, cover important and entrustable aspects of professional practice. The consensus framework of EPAs that emerges from the process may lose the critical alignment with professional practice that is targeted.

Evaluating EPA construction requires clearly defined quality metrics for EPAs as well as a means of measuring them. Various publications have laid out the quality standards and common pitfalls in EPA construction that should anchor this work.^{2,4,5} One approach is designating two to three core team members with EPA expertise as responsible for evaluating the quality of drafted EPAs and making revisions to improve their formulation. A more rigorous approach involves applying a rubric designed to measure EPA quality standards, such as the EQual rubric tool and the QUEPA instrument.^{5,6} It is important to note that these rubrics were not created to rubber-stamp high-quality EPAs; instead, they provide focused feedback on the attributes of each EPA to the development team identifying areas for improvement in problematically constructed EPAs. Additionally, using a rubric may identify more systematic problems in a project. If data collected from evaluating the EPAs shows pervasive quality problems across multiple EPAs, this may be flagging faculty development needs around EPA definitions and construction.

A deeper dive into the EQual rubric

As the EPAs are the final product and purpose of the project, it is hard to overemphasize the importance of attending to the quality of each EPA being produced in any EPA development project. Assessing individual EPA quality requires a clear understanding of their purpose in the curriculum, and the quality standards for EPAs that support this purpose.^{2,4} At a foundational level, the purpose of EPAs is ‘operationalizing competencies to facilitate reliable assessment, which predicts future performance at a defined professional standard,’ and the construction of EPAs must support this.^{1,5}

In constructing the EQual rubric, Taylor et al. identified the defining features and important misconceptions of EPAs that have been published in the peer-reviewed literature.^{2,4} Each feature and misconception was then reviewed to ensure its relevance to the above-described purpose of EPAs. All eight defining features clearly supported one or more elements of this purpose, and the six identified misconceptions represented potential quality threats. This provided the 14 quality standards that would be used in constructing the EQual rubric (see Table 11.2).⁵ These standards are naturally organized into three domains: items describing EPAs as discrete units of work (items 1–6); items related to entrustment and professional standards (items 7–10); and items that enhance application in curriculum (items 11–14). Criterion-referenced, five-point numeric rating scales were then developed for all 14 items to enable scoring.

Table 11.2: Explanation of items in the EQual rubric.

Domain	EQual rubric item	Explanation
Units of work	1. This EPA has a clearly defined beginning and end	The EPA is a discrete task that has a clear beginning and end, and is completed within a discrete period of time
	2. This EPA is independently executable to achieve a defined clinical outcome	The activity does require performance of other related tasks to produce its clinical outcome
	3. This EPA is specific and focused	The task is clearly recognizable to those in the profession with the steps required to perform broadly agreed upon
	4. This EPA is observable in process	The process to performing the EPA can be observed (or directly interrogated) by a clinical supervisor
	5. This EPA is measurable in outcome	Completion of the EPA by a learner produces an outcome that can be assessed (e.g., completion of a consultation allows for a judgment of ‘well done’ or ‘not well done’)
	6. This EPA is clearly distinguished from other EPAs in the framework	The EPA does not describe clinical work that is also captured in other proposed EPAs
Entrustment and professional standards	7. This EPA describes work that is essential and important to the profession	EPAs should identify the core work that all individuals in the profession should be certified to perform at the end of training
	8. Performing this EPA leads to recognized output or outcome of labour	Performing the EPA should lead to a tangible product that advances clinical care (e.g., a patient treatment plan produced after a medical consultation)
	9. The performance of this EPA in clinical practice is restricted to qualified personnel	The EPA belongs within the scope of practice of the profession and is not performed by unqualified individuals without appropriate supervision
	10. This EPA addresses professional work that is suitable for entrustment	These are tasks that require specific expertise to perform safely and competently that society has entrusted to the profession
Curriculum application	11. This EPA requires the application of knowledge, skills, and/or attitudes (KSAs) acquired through training	To perform the EPA, learners must acquire knowledge, skills, and/or attitudes that inform the competencies required of the task
	12. This EPA involves application and integration of multiple domains of competence	The EPA does not reflect abilities from a single domain of competence (e.g., communication), but is a task that requires integration of multiple competencies to perform (e.g., history and physical examination)
	13. The EPA title describes a task, not qualities or competencies of a learner	The EPA should be work done in caring for a patient (e.g., taking a history), not an ability possessed by a learner (e.g., patient-centred communication skills)
	14. This EPA describes a task and avoids adjectives (or adverbs) that refer to proficiency	Adjectives may be included in EPAs (e.g., ‘...common medical presentations’) but should not reference effectiveness of the learner’s performance (e.g., ‘...obtaining an appropriate history and physical exam’)

More than offering a single, summative score, grouping the 14 items into the three domains allows for more nuanced and formative evaluation of EPAs. Ratings within domains provide important information about areas for revision and improvement. Well-constructed EPAs may have items with lower scores that are offset by items with higher scores. When lower ratings occur, it is up to users to determine their significance—does a low score identify a need for revision, or does it represent acceptable variation in scoring within an otherwise acceptably scored domain? Application of EQual was not intended to provide definitive, summative judgment on EPAs; instead, it was meant to direct revisions and improvements to EPAs in an evidence-based manner. The overall cutscore for the rubric (an average item rating of 4.07, which was determined using a modified Angoff approach) is used primarily as a screening test to effectively distinguish EPAs that need major revisions from those that may need only minor revision or no revisions at all.^{5,7}

This leads to the question of when to use EQual in EPA development. Although there is some value in using it at or near the end of EPA development (in advance of formally adopting a framework of EPAs), its formative value advocates for employing it early in the development process. In addition, training project team members to use the EQual rubric helps build a shared understanding around the standards for EPAs, something that is important to establish early. Ideally, EQual should be used early enough in the process that it can impact the quality of material prior to entering consensus methods such as Delphi. This needs to be balanced against the time required for its use; if there are more than 30 or 40 preliminary EPAs under consideration at this stage, the time required to apply the rubric may become prohibitive. In many cases, it may be beneficial for project leaders to revise preliminary EPAs down to a reasonable number prior to deploying EQual.

There are several principles to using EQual that are important to ensure you get the most out of it. A minimum of four EPA experts should rate the EPAs with the rubric. Studies in both undergraduate and postgraduate medical education suggest that having four expert raters provides excellent reliability.^{5,7} That being said, involving five or six experts will increase the number of perspectives and enrich the feedback being collected on the EPAs. Keep in mind, applying EQual is more about collecting high-quality feedback to direct revisions than it is about getting a reliable quantitative judgment on EPA quality.

Defining what constitutes expertise for those recruited for the EQual exercise is also important. These individuals should have a background in CBME with expertise in EPAs. Evidence of expertise might come from attendance at courses and workshops, or scholarly publications in the field. Even with established experts, it is important to provide training on using the EQual rubric to ensure a shared mental model for the task. The EQual training video was designed for this purpose and is readily available on YouTube.⁸ See Chapter 25 for an EQual Rubric tool.

Finally, it is important to have EPAs developed to a point that they are suitable for evaluation by the rubric. It is not expected that EPAs will be fully elaborated prior to using EQual; however, having only a title is inadequate. At a minimum, EPAs descriptions must include the title and the specifications and limitations that clearly describe the steps to performing the EPA. These descriptions should be provided to those evaluating the EPAs using EQual.

Results from using EQual (both comments and quantitative evaluation) should be reviewed by the development team. These can then direct revisions to EPAs. Teams may also consider eliminating EPAs that are fundamentally flawed and not amenable to revision to an acceptable EPA.

Evaluation strategies for EPA revision and refinement

Once an initial set of EPAs has been proposed and the construction of the EPAs evaluated and deemed appropriate, consensus methods are then typically used to direct further revision and determine suitability for inclusion in the final EPA framework. Delphi techniques are commonly used survey methods for this purpose. As described in Chapter 9, Delphi iteratively builds

consensus for proposed EPAs by surveying participants through multiple rounds, revising and refining the EPA descriptions based on feedback collected after each round. Survey rounds repeat until a designated measure of agreement is achieved for the EPAs (or no further progress toward consensus is reasonably expected with additional rounds). In Delphi, the development method and the evaluation method are integrated.

Delphi can be used early in the process in a more formative manner to direct EPA development and revision.⁹ The method can also be used later in a more summative manner to solicit broad, frontline clinician feedback on a proposed framework of EPAs and to identify gaps (in this context it is often referred to as a reactor panel).¹⁰ Whenever Delphi is used, it is important to monitor for known problems that can emerge with it. There are several publications that outline approaches to monitoring for problems in use of Delphi for this context.^{1,11,12}

It can be appreciated that this phase of the project requires a transition in evaluation focus. Early on, the focus is formative. However, by the end of this phase it becomes more summative, with the aim of having a defensible framework of EPAs ready for field testing. This requires a shift in evaluation from focusing on quality standards directing revisions at the beginning to considering acceptability to end users and regulatory bodies by the end. Delphi methods can often provide answers to both of these questions, depending on those engaged as participants in the Delphi.

Evaluating your end product

Can these newly developed or revised EPAs achieve the assessment and quality assurance sought by regulators? Can they support the curricular purposes required by programs and accreditors? The answers to these questions are hypothetical until real-world experience is gained. During piloting, gather information from key groups including programs, frontline clinicians, and learners, to explore authentic experiences of end users. In doing so, take the time to investigate challenges experienced and how they relate to (a) implementation, (b) the EPA descriptions, and (c) curriculum. Because these areas are not independent but highly interdependent, challenges experienced during the pilot will likely relate to two or three of these areas. This exploration can provide crucial information for final adaptations and to establish feasibility for meeting the needs of accreditors and regulators.

In doing this work, it is crucial to try to discriminate between challenges associated with introducing change and challenges stemming from the quality of the EPA framework, a task that can be difficult to near impossible. Unfortunately, leaders introducing EPAs have at times been dismissive of concerns raised during implementation, often labeling concerns as operational problems and not issues with the EPAs or curricular design.¹³ As projects proceed through pilot testing, evaluation should err on the side of acknowledging and responding to concerns raised about the EPAs, even when concerns might otherwise be classified as implementation problems. Chapter 22 discusses change management strategies when implementing EPAs.

Monitoring after implementation

Evaluation does not end with adoption and implementation. As implementation of EPAs proceeds to scale, unexpected challenges and new learnings inevitably emerge. Stakeholders can report on the impact of the implementation—are learners better trained? Are patients better served? Have there been other unexpected consequences to the new framework? Some of these evaluation questions may take years to answer. Prepare for this with a coordinated plan to collect and review feedback at regular intervals, allowing time (e.g., five years) to monitor implementation prior to considering major revisions. There is no one prescription for this. Instead, leaders should leverage

established program evaluation infrastructure that already exists at various program sites and coordinate regular reviews of progress and challenges. For larger projects, such as with national-level curricular change, connecting different program sites to develop a network for systematically collecting data is important and can identify system-wide problems early. An iterative and deliberative approach to monitoring after evaluation, tailored to the scope and nature of the project, is fundamental to long-term success.

Conclusion

Ensuring high-quality EPAs is essential to success when introducing an EPA-based curriculum. The evaluation strategies discussed in this chapter can help teams tasked with developing EPAs to achieve this goal. There is no one-size-fits-all strategy to evaluation, and leaders must prioritize evaluation methods most relevant to their projects and in consideration of resource constraints. Thoughtful evaluation planning will not only improve the quality of the EPAs and frameworks produced but can also save time and resources.

Competing interests

The authors declare that they have no competing interests.

References

1. Taylor D, Park YS, Smith C, ten Cate O, Tekian A. Constructing approaches to entrustable professional activity development that deliver valid descriptions of professional practice. *TLM*. 2020;2(3):1–9. DOI: <https://doi.org/10.1080/10401334.2020.1784740>
2. 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: <https://doi.org/10.1097/acm.0b013e31805559c7>
3. Fitzpatrick J, Worthen B, Sanders J. *Program Evaluation Alternative Approaches and Practical Guidelines*. 4th ed. Pearson Education Inc.; 2011.
4. 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: <https://doi.org/10.3109/0142159x.2015.1060308>
5. Taylor DR, Park YS, Egan R, et al. EQual, a novel rubric to evaluate entrustable professional activities for quality and structure. *Acad Med*. 2017;92(11S):S110–S117. DOI: <https://doi.org/10.1097/acm.0000000000001908>
6. Post JA, Wittich CM, Thomas KG, et al. Rating the quality of entrustable professional activities: content validation and associations with the clinical context. *J Gen Intern Med*. 2016;31(5):518–523. DOI: <https://doi.org/10.1007/s11606-016-3611-8>
7. Meyer EG, Taylor DR, Uijtdehaage S, Durning SJ. EQual rubric evaluation of the Association of American Medical Colleges' core entrustable professional activities for entering residency. *Acad Med*. 2020;95(11):1755–1762. DOI: <https://doi.org/10.1097/acm.0000000000003504>
8. Taylor DR, Park YS, Tekian A. *Queen's EPA Quality (EQual) Rubric Training*. Vol 17:41.; 2016. Accessed March 26, 2024. <https://youtu.be/yQZuWdzkQKM>
9. Taylor DR, Park YS, Smith CA, Karpinski J, Coke W, Tekian A. Creating entrustable professional activities to assess internal medicine residents in training: a mixed-methods approach. *Ann Intern Med*. 2018;168(10):724–729. DOI: <https://doi.org/10.7326/m17-1680>

10. Englander R, Flynn T, Call S, et al. Toward defining the foundation of the MD degree. *Acad Med*. 2016;91(10):1352–1358. DOI: <https://doi.org/10.1097/acm.0000000000001204>
11. Powell C. The Delphi technique: myths and realities. *J Adv Nurs*. 2003;41(4):376–382. DOI: <https://doi.org/10.1046/j.1365-2648.2003.02537.x>
12. Keeney S, Hasson F, McKenna H. *The Delphi Technique in Nursing and Health Research*. 1st ed. Wiley-Blackwell; 2011. DOI: <https://doi.org/10.1002/9781444392029>
13. Boyd VA, Whitehead CR, Thille P, Ginsburg S, Brydges R, Kuper A. Competency-based medical education: the discourse of infallibility. *Med Educ*. 2017;32(8):638. DOI: <https://doi.org/10.1111/medu.13467>