Critical Thoughts About the Core Entrustable Professional Activities in Undergraduate Medical Education
Edward Krupat, PhD

Abstract
The Core Entrustable Professional Activities for Entering Residency (Core EPAs) have taken a strong hold on undergraduate medical education (UME). This Perspective questions their value added and considers the utility of the Core EPAs along two separate dimensions: (1) the ways they change the content and focus of the goals of UME; and (2) the extent to which entrustable professional activity (EPA)–based assessment conforms to basic principles of measurement theory as practiced in the social sciences. Concerning content and focus, the author asks whether the 13 Core EPAs frame UME too narrowly, putting competencies into the background and overlooking certain aspirational, but important and measurable, objectives of UME. The author also discusses the unevenness of EPAs in terms of their breadth and their developmental status as core activities. Regarding measurement and assessment, the author raises concerns that the EPA metric introduces layers of inference that may cause distortions and hinder accuracy and rater agreement. In addition, the use of weak anchors and multidimensional scales is also of concern. The author concludes with a proposal for reframing the Core EPAs and Accreditation Council for Graduate Medical Education competencies into broadly defined sets of behaviors, referred to as "Tasks of Medicine," and calls for the development of a systematic and longitudinal research agenda. The author asserts that "slowing down when you should" applies to medical education as well as patient care, and calls for a reevaluation of the Core EPAs before further commitment to them.

Editor’s Note: An Invited Commentary by E.S. Holmboe appears on pages XXX–XXX.

With all due respect to the experienced and thoughtful medical educators who have introduced, endorsed, and led the movement in undergraduate medical education (UME) toward the Core Entrustable Professional Activities for Entering Residency (Core EPAs),1–5 I want to say that I do not like them (the Core EPAs, not the educators). I recognize that my reservations about the growing use and adoption of the Core EPAs may be applauded by some and reviled by others, but I write urging more public debate, to throw light, not just heat, on the pros and cons of this movement.

Let us first take a step back to provide some context. The world of competency-based medical education (CBME) is constantly evolving. Following the initial focus on competencies (domains in which health professionals are expected to demonstrate knowledge and skills), as defined by the Accreditation Council for Graduate Medical Education (ACGME),6,7 the concept of milestones8,10 (observable markers along the path to competence) was proposed. Concerned that the ACGME’s six core competencies did not specify concretely the tasks, actions, or responsibilities required of a physician, ten Cate proposed the concept of entrustable professional activities (EPAs) in the context of graduate medical education (GME).11,12 These EPAs represented a pragmatic list of activities—a core set of tasks or responsibilities that all residents could be trusted to do independently by the time they completed their training.

With the growing acceptance of EPAs in GME, the Association of American Medical Colleges (AAMC) recognized a “performance gap at the transition point between medical school and residency.”13 The AAMC convened a panel of respected educators to identify those activities that might be expected of all graduating MDs, with the goal of having GME and UME work within a common framework. Aiming to ensure that all medical school graduates would be entrustable to perform the tasks of a new resident on Day 1,14 the AAMC adopted a set of 13 Core EPAs and put together an infrastructure to study them further.15,16

Considerable organizational resources have gone into piloting and testing the Core EPAs.15 Some educators have described EPAs as important and paradigm shifting;16 however, a growing number have expressed reservations about CBME in general and EPAs in particular.17–22 One set of observers asks us to slow “the headlong rush to competencies, milestones and CORE-EPAs … before even more time, effort, frustration and resources are invested.”17 In the context of errors in diagnostic reasoning, Moulton and colleagues23,24 have talked about “slowing down when you should.” Perhaps now is the right time to pause, take a deep breath, and ask ourselves whether we are taking a major step in the right direction rather than a step back or to the side.

My informal review of published critiques of the EPAs suggests that they vary in two ways. Some question the content or focus of the EPAs, arguing that we are measuring the wrong things, measuring the measurable rather than the important. Others raise issues of measurement and method, noting problems in making valid and reliable assessments. Most often, however, these
critiques have confounded problems of content and method, making it difficult to identify and isolate key issues of different types. In this Perspective, I will try to address these two dimensions independently.

**Concerns About Content and Focus**

Rather than focusing on domains of competence (e.g., medical knowledge) or characteristics associated with individuals (e.g., professionalism), the Core EPAs constitute a list of 13 ambulatory and hospital-based “activities” (discrete units of professional behavior expressed as workplace-based tasks) that all medical students should be able to accomplish by their graduation (and, consequently, on Day 1 of internship). Tekian has criticized the Core EPAs on several counts, noting that three of them (e.g., collaborating in an interprofessional team) are not actually discrete, single-encounter-based medical tasks, while others sound more like educational objectives than tasks (e.g., forming clinical questions and retrieving evidence).

In addition, the Core EPAs seem to me like an uneven lot, with some extremely broad and others very specific. I can imagine that the expert panel that decided on these EPAs was composed of “lumpers” (who preferred to group things into broad categories) and “splitters” (who preferred to generate finer categories), and that each group had its way on some things. For example, gathering a history and performing a physical exam (EPA 1) are typically thought of as two core tasks of the student–physician, yet these two activities were combined. On the other hand, written documentation (EPA 5) and oral presentation (EPA 6) of a clinical encounter could have been included as a single EPA, “summarizing a clinical encounter,” but the panel members chose to list these as two separate activities.

Not only are the Core EPAs uneven in size and scope but they are also uneven from a developmental perspective. Some of them are part of the bread-and-butter of being a physician, whereas others appear to be quite advanced. I was recently on a committee that considered how many of the Core EPAs would be relevant for assessing students in their preclerkship versus core clerkship years. Some argued that the assessment form should omit certain EPAs for students in their early years or include a “Does not apply” category. For instance, learning about handovers (EPA 8) and recognizing patients requiring urgent care (EPA 10) seem to be the stuff of capstone courses and preinternship boot camps rather than “core” activities that medical schools ought to be teaching throughout the curriculum. Experienced GME educators have told me that identifying system failures (EPA 13) seems aspirational for their second-year residents, let alone graduating medical students.

I have a far more fundamental concern about the Core EPAs, however. Advocates of this approach note that the Core EPAs exist to complement, not replace, graduation competencies, and they have provided extensive discussions of the relationship between the two. Their message is that the issue of competencies versus EPAs should not be seen as something that is either–or.

In principle this is fine. In practice, however, this ignores one of the most basic and universally recognized principles of human perception: figure–ground relationships. When we are considering two “objects,” one (figure) engages our attention and is focused upon. The other (ground) recedes into the distance and is taken for granted; it serves as context, not as the object of interest. In suggesting that EPAs serve as the benchmarks for medical education internationally, ten Cate has asserted that the movement of the EPA concept from GME to UME “is in fact a redefinition of what a medical physician is, in terms of what he or she should be allowed to do.” In effect, this suggests that EPAs should serve as figure while other aspects of physicianship recede to ground. But is this a step in the right direction? Is it to our advantage to make the Core EPAs figure, while relegating to ground professionalism, the ability to communicate with patients, and the ability to deliver quality patient care regardless of race/ethnicity, gender, or sexual orientation?

EPA advocates seem to bear no ill will toward the competencies. Even so, competencies are likely to become second-class citizens in the curricular map if they are not at the forefront of our educational objectives and assessments. Medical schools that adopt the Core EPAs will not be rewriting their mission statements, but I do not think it is an exaggeration to say that if formative and summative assessments come to be focused on students’ ability to perform these 13 EPAs, this is what both faculty and students will care about most. Anyone familiar with the phrase “Assessment drives learning” can hardly wonder what will become primary in the curriculum and what secondary; I have little doubt as to which will become figure and which ground.

I am also concerned that there is nothing aspirational in the Core EPAs and nothing in their language that implies that medicine is a calling with a core set of professional values. Others before me have expressed concerns that a reductionist, concrete-task-based focus will affect the identity formation of young physicians by not focusing on issues of interconnectedness and the broader meaning of being a physician. Taking this one step further, I ask to what extent the Core EPAs also say something symbolically to patients and society. An EPA-focused education will produce physicians who know how to satisfy the technical requirements of care and to fulfill the tasks that enable a hospital to run efficiently. However, to produce inspired professionals, within our educational objectives and assessments we need to find a way to feature the broader values and orientations—as well as the activities—that are part of great doctoring.

Further, how is forming a caring relationship with a patient any less a discrete “unit of behavior” that can be observed and assessed in a patient encounter than forming clinical questions and retrieving evidence (EPA 7)? Although it might seem at first as if operationalizing elements of caring between physician and patient is challenging, several frameworks exist, such as the Four Habits Model, that can enable us to assess entrustment in this professional activity.

Finally, if a prime motivator in the Core EPA movement is to bridge the gap between UME and GME in terms of preparation for Day 1 of internship, then it would follow logically that this EPA-based information is what program
directors want to know about their applicants to make residency selection decisions. The evidence indicates, however, that program directors are most interested in potential residents' personal qualities (e.g., their ability to work with others, reliability, and honesty—elements of professionalism), rather than whether they are entrustable on this or that Core EPA.33,34

What sort of gap has been bridged when the things we are assessing for the sake of the residencies and the information program directors want are distinctly different?

**Concerns About Measurement and Assessment**

The second dimension of my concerns involves the assessment process itself, as the EPA approach fundamentally changes themetric to be used in assessment, substituting the concept of judgment via entrustment for a variety of other approaches used in pre-EPA assessment. On the positive side, the concept of trust is intuitively appealing. Most physicians are familiar and comfortable with thinking along a dimension that relates to how well they can trust their trainees, at all levels, to accomplish certain goals and tasks. Certainly, the potential buy-in this generates among physicians should not be underestimated. Nonetheless, I believe that many aspects of EPA assessment are not sufficiently informed by measurement theory in the social sciences, and that many aspects of the process fly in the face of established principles of good measurement practice.35,36

Despite our desire for objectivity, until we invent a thermometer to take the temperature of people's behaviors and characteristics, we are left with a subjective process of assessment. Given this, one of our main goals is to find an approach by which we can increase interrater agreement (i.e., reliability). To do so, the golden rule in measurement theory is to keep it simple: See a behavior and judge it according to a single evaluative scale, thereby making a rating. What should be minimized at all costs is making judgments in which layers of inference are placed between the observation and the judgment.37

The principle is the same as that of the children's game of "telephone." The more removed the eventual judgment is from the raw data, the more opportunity there is for distortion and inaccuracy.

To see how assessment via entrustment fares according to this criterion, let us enumerate the layers. First, we observe the behavior, and then we determine for ourselves whether it has been performed competently—which, if stopped here, would be a reasonable rating system. Next, we consider trustworthiness and ask ourselves how much trust the student inspires in us to perform a similar behavior (on a similar patient with a similar problem), while simultaneously considering how closely this student would need to be supervised to do so, all the time acknowledging the limits of making generalizations because of context specificity.38,39 Adding more layers, research suggests that entrustability judgments ultimately go beyond knowledge and skillfulness to include assessments of characteristics such as truthfulness, conscientiousness, and discernment of limitations.40,41

Taking all these considerations together, assessment via entrustment hardly appears to be a system that would reduce layers of inference and increase accuracy and agreement among raters.

Another rule of thumb in good measurement is to ensure that the end points, or anchors,42,43 for any given scale are clearly recognized, defined, and understood by all relevant parties in exactly the same way. This is important yet extremely difficult. It is why those knowledgeable about careful measurement are wary of weakly anchored scales that use designations such as "poor" and "excellent." In GME, the entrustment scale is anchored on the positive end by wording indicating that the resident can be entrusted to perform a task completely unsupervised. This anchor, stated in absolute terms as the total absence of supervision, is compelling. Using such an anchor summatively, a program certifies that its graduates are entrusted to practice in complete independence, to perform the EPAs with no oversight or supervision. In UME, however, the anchor on the positive end indicates that the student can be entrusted to practice a given EPA under indirect supervision. This anchor is far less grounded; it is far less well defined. How indirect is indirect supervision? How much autonomy should we grant a medical student under conditions of indirect supervision?

In attempting to be more creative about Core EPA entrustment scales, others have proposed different anchors, such as defining the positive end point in terms of whether the person could be "allowed to supervise others in practice." As intuitively compelling as this example seems, it violates another cardinal rule of good measurement, which is that a scale should vary along only one dimension at a time. Could we not imagine a trainee who is quite capable of performing an EPA yet whom we would not want supervising others in that EPA? Good measurement practice demands that scales be defined unidimensionally so as to avoid confounding the rating criteria, but this proposal runs counter to that principle.

**Other Problems Inherent in Working With the Core EPAs**

In addition to the content- and measurement-related issues raised above, several logical inconsistencies seem to be largely unrecognized and unresolved in the Core EPA movement. First, because these EPAs were meant to bridge the gap between UME and GME, we can place the UME and GME rating categories side by side to determine the consistency with which they are used in the two settings. For instance, at the low end of GME entrustment, one often finds "allowed to practice EPA only under proactive, full supervision."44 Yet how could an intern be rated as requiring proactive, full supervision when in order to graduate medical school this same person had to be certified as able to practice with indirect, or minimal, supervision?

A related question has to do with summative decisions and timing. Everything we know about the variability in human behavior tells us that even among the talented individuals who enter medical school or residency, differences will exist in their ability to meet high standards and the pace at which they do so; some will progress quickly, while others will need additional time.45 Yet how is it that almost every member of every medical school class or residency cohort would come to meet a demanding and carefully assessed set of EPA standards at the same moment—at the point of their final summative
assessments, just as their graduation date is approaching?

One possible answer, which we would all prefer to reject, is that the standards are sufficiently low that virtually everyone, even poorly performing students, can meet them on time. Offering an even more skeptical thought, Witeles and Verghese\textsuperscript{4} have speculated whether overworked assessors may sometimes be checking boxes for where students are supposed to be rather than attempting to provide assessments that are accurate and meaningful. I would hope that neither of these answers explains this paradox.

Another explanation, deriving from social cognitive theory, is more charitable to the standards and assessors: Standards implicitly differ when seen in the differing contexts of UME and GME. The \textit{very same behavior} that was rated as entrustable against the responsibilities of UME studenthood may appear to be preentrustable against the standards of residency where house staff hold far greater responsibility for patients' well-being. Referring to \textit{contrast effects}, social psychologists suggest that raters may \textit{unwittingly} make differing judgments as the result of judging behavior against different standards in different settings. For instance, a football player who was considered a trustworthy and seasoned leader of his college team as a senior is likely to be seen as an immature and unreliable rookie when he moves up a step from the college to the professional ranks, even though his level of ability has not diminished. In a similar manner, the same behavior that inspired trust in a fourth-year medical student may seem less trustworthy for an intern. Two other closely related concepts, \textit{expectancy effects} and \textit{self-fulfilling propheses}, posit that when we hold specific expectations for behavior (e.g., that trustworthiness is greater with increased experience and training), these may influence us, below the level of awareness, to make judgments in the direction of our expectations.\textsuperscript{46,47}

In biomedical research, we insist on double-blind studies, and an R01 application proposing research with unblinded raters is not likely be funded by the National Institutes of Health. But in clinical skills assessments made in vivo, raters are almost never blinded to the experience level of the person being rated. As a result, raters may be subject to various nonconscious, automatic biases, whatever their level of faculty development or motivation. In effect, the “ruler” we are using may simply be more elastic than we realize, and EPA-based judgments do little if anything to overcome problems of ratings bias.

Conclusions and Recommendations

When sharing family recipes, many of our grandmothers instructed us to add a pinch of salt or a sprinkle of pepper. If I believed that the concept of entrustment had turned a pinch or sprinkle into something objective such as ounces or milligrams, I would be overjoyed. However, for all the effort expended by so many talented people, I do not believe that the adoption of an EPA-based system has increased our ability to make reliable ratings or valid judgments, to provide students with useful formative feedback, or to make defensible summative decisions about progression.

As for content, I believe that the movement toward EPAs has made us focus too much on what Grandma knows how to cook, not what is good for us.\textsuperscript{10} The Core EPAs direct too much of our attention toward the mundane and technical skills that allow hospital teams to function smoothly, thereby distracting us from those roles and behaviors of physicians that are more abstract in concept, yet could be operationalized behaviorally if only we worked harder at doing so. Methodologically, although physicians may be comfortable using trust as their frame of reference, assessment via EPAs does not decrease inference or subjectivity. While EPA-based measurement ought to build on the strengths of measurement theory, it violates several key principles of good measurement, thereby introducing noise and facilitating bias, however unintended.

Rather than merely criticize, it is important to propose reasonable alternatives and put forward recommendations. So let me offer a few modest proposals concerning both content and assessment. First, while some have argued that EPAs and competencies can coexist side by side,\textsuperscript{48} recalling the principle of figure and ground, I believe that educators are likely to focus on one or the other—to the extent that they are seen as separate entities. Therefore, I would like to propose a single hybrid model—in the spirit of the CanMEDS roles\textsuperscript{4} and encompassing both the Core EPAs\textsuperscript{11} and the ACGME competencies—that broadens our notions of a “professional activity.”

Concerning assessment, although a key objective of the Core EPAs was to bridge the UME–GME gap, it still seems as if undergraduate and graduate medical educators are working in parallel. A great deal of research has been sponsored by the AAMC for medical schools to pilot the Core EPAs,\textsuperscript{15} and parallel efforts exist in GME.\textsuperscript{5,6,8} However, because different frames of reference, standards, and expectations may exist for clerks, interns, and residents, identical behaviors may be judged differently depending on their context. Therefore, we need to develop a \textit{single set of descriptors} that covers the \textit{full range} from beginner to master, from third-year student to chief resident to 20-year practitioner. If we can define and provide examples of the full spectrum of expected behaviors for each TOM, we will be far less subject to contrast effects and self-fulfilling propheses in both UME and GME assessments.\textsuperscript{52,53}

In addition, I call for greater efforts to create a comprehensive UME–GME research agenda. This would begin with more longitudinal research that follows students through medical school and residency, something that few researchers have done systematically. Of course, this would mean following students not only over time but also across institutions; however, if we want to understand how behavioral competence grows developmentally in order to set appropriate standards, then we must follow learners over the \textit{full course} of their training.

Another critical need is to determine the extent to which nonblinded assessments affect our ratings. With appropriate ethical and privacy protections ensured, it should not be difficult to video record students...
List 1

Tasks of Medicine, or What Great Physicians Do: A Hybrid Model Attempting to Integrate Behavioral Elements of the Core EPAs With the ACGME Competencies

On the first day of residency, an entrustable physician can, with minimal instruction or assistance
By the completion of residency, an entrustable physician can, with no instruction or assistance

- Collect and apply information in the care of patients
  - Take a complete and accurate history
  - Examine patients
  - Order tests appropriately
  - Retrieve evidence from appropriate sources

- Synthesize and utilize information
  - Make diagnoses
  - Develop treatment plans
  - Recognize and initiate urgent care for those patients who require it

- Work effectively with others
  - Give oral reports and document encounters in patient records
  - Give and receive handovers effectively
  - Work effectively with other health professionals
  - Develop caring and trusting relationships with patients and families
  - Obtain informed consent from patients and families

- Practice medicine demonstrating awareness of the larger context
  - Provide effective care to diverse populations regardless of race, ethnicity, income, etc.
  - Identify and attempt to address system failures

- Demonstrate responsibility, self-awareness, and desire for improvement
  - Make efforts at self-improvement based on feedback
  - Demonstrate reliability and honesty in the course of ethical practice
  - Recognize strengths and limitations in the diagnosis and treatment of patients

Abbreviations: Core EPAs indicates Core Entrustable Professional Activities for Entering Residency; ACGME, Accreditation Council for Graduate Medical Education.

and trainees who are at different levels of training and who display varying levels of competence (for one or more TOMs) in a range of workplace-based situations, and to ask assessors (some blinded, others with knowledge of the trainee’s status) to rate them. This would allow us to design a set of randomized controlled trials that would (1) determine empirically the extent to which assessments are affected by knowledge of the trainee’s status and (2) guide us toward measurement solutions (e.g., testing different scales with different anchors) that avoid this source of potential bias and that have strong discriminatory ability and maximize interrater agreement.

The EPA movement is a train that has not only left the station but has gained considerable momentum. I assert that in an attempt to address several important problems in competency assessment, the Core EPAs have created others, and it is unclear the extent to which we have a net gain or a clear sense of value added. There is great virtue in “slowing down when you should,”23 thereby avoiding the trap of continuing to commit additional resources to an enterprise merely to justify the effort already expended. We need to consider alternative ways of bridging the UME–GME divide, and we need to reevaluate the content and assessment methods of EPAs in general and the Core EPAs in particular. This will allow us to decide whether this approach should serve as the focus of attention of the medical education community and the metric for assessment to which we want to commit long-term.

Funding/Support: None reported.

Other disclosures: None reported.

Ethical approval: Reported as not applicable.

References


Perspective


21 Gingerich A. What if the “trust” in entrustable were a social judgement? Med Educ. 2015;49:750–752.


37 Norcini J. President and chief executive officer, Foundation for the Advancement of Medical Education. 2017. Personal communication with E. Krupat.


