Facilitator: Tanya Horsley, PhD
Discussant: Karen Hauer, MD

EQual, a Novel Rubric to Evaluate Entrustable Professional Activities for Quality and Structure

David R. Taylor, MD, MHPE, Yoon Soo Park, PhD, Rylan Egan, PhD, Ming-Ka Chan, MD, MHPE, Jolanta Karpinski, MD, MEd, Claire Touchie, MD, MHPE, Linda S. Snell, MD, MHPE, and Ara Tekian, PhD, MHPE

Abstract

Purpose
Entrustable professional activities (EPAs) have become a cornerstone of assessment in competency-based medical education (CBME). Increasingly, EPAs are being adopted that do not conform to EPA standards. This study aimed to develop and validate a scoring rubric to evaluate EPAs for alignment with their purpose, and to identify substandard EPAs.

Method
The EQual rubric was developed and revised by a team of education scholars with expertise in EPAs. It was then applied by four residency program directors/CBME leads (PDs) and four nonclinician support staff to 31 stage-specific EPAs developed for internal medicine in the Royal College of Physicians and Surgeons of Canada's Competency by Design framework. Results were analyzed using a generalizability study to evaluate overall reliability, with the EPAs as the object of measurement. Item-level analysis was performed to determine reliability and discrimination value for each item. Scores from the PDs were also compared with decisions about revisions made independently by the education scholars group.

Results
The EQual rubric demonstrated high reliability in the G-study with a phi-coefficient of 0.84 when applied by the PDs, and moderate reliability when applied by the support staff at 0.67. Item-level analysis identified three items that performed poorly with low item discrimination and low interrater reliability indices. Scores from support staff only moderately correlated with PDs. Using the preestablished cut score, PDs identified 9 of 10 EPAs deemed to require major revision.

Conclusions
EQual rubric scores reliably measured alignment of EPAs with literature-described standards. Further, its application accurately identified EPAs requiring major revisions.

Entrustable professional activities (EPAs) are becoming the backbone of assessment in competency-based medical education (CBME).[1–6] As units of work, they offer the advantage of being directly observable and measurable, integrating knowledge, skills, and attitudes into authentic professional tasks.[7,8] For training programs in CBME, EPAs aim to strengthen the assurance in the outcome of training by operationalizing competencies to enable assessment that predicts future performance at an acceptable standard.

Please see the end of this article for information about the authors.

Correspondence should be addressed to David R. Taylor, Queen's University, Department of Medicine, Etherington Hall, 94 Stuart St., Kingston, Ontario, K7L 3N6 Canada; telephone: (613) 533-2623; e-mail: taylord@queensu.ca.

Acad Med. 2017;92:S110–S117.
doi: 10.1097/ACM.0000000000001908

EPAs have been carefully described in order that the elements of their definition collectively work toward this goal.[4,9] Demonstrating that an EPA meets the requirements of this definition is therefore essential to generating EPAs that meet their purpose in facilitating reliable assessment of professional work that is predictive of future performance.

Despite the clear link between the definition and the goal, there is diversity in published EPAs on the degree to which they meet the defining qualities of EPAs.[4,10] This problem persists despite the use of various consensus-building and evaluation approaches in developing EPAs.[3–5,11,12] There is a clear need for tools to assist in evaluating EPA quality to ensure that the final product holds to the standards required.

Post et al[13] developed and evaluated the only tool (Quality of Entrustable Professional Activities [QUEPA]) to measure the quality of EPAs that the authors are aware of. It was thoughtfully developed with attention to a variety of elements of EPAs, and produced scores with good interrater reliability. However, the application of the tool has several limitations. First, the development team defined seven essential domains of quality for EPAs; although their process began with a review of the literature, the final domains did not completely align with the defining qualities of EPAs as described.[4,7] Second, the QUEPA tool uses normative scales for each item without descriptive anchors. EPA development demands that EPAs meet specified standards and therefore should use criterion-based evaluation. Third, the data generated in validating the tool came from users who were part of the tool development team. As developers, they would have had an inherent shared mental model; other users may not benefit from this and would require user-

training resources, which have not been defined. Fourth, there was no attention to establishing cut scores for items to define what is acceptable. Although sensible and carefully developed, these limitations can hinder the QUEPA tool's usability in a broader context.

Our study describes the development and validation of a tool for measuring EPA quality (EQual: Queen's EPA Quality Rubric; see Appendixes 1 and 2) that was criterion based and was tied closely to the defining features of EPAs as well as misconceptions of EPAs as described in the peer-reviewed literature. In addition, we wanted to evaluate whether the tool could be used by nonclinician support staff without clinical experience; this would broaden its potential use in scholarly projects on EPAs. As part of our implementation process, we defined cut scores to aid EPA developers in applying our tool and developed an online video resource for training users.

Method

Ethics approval for this study was obtained from the Research Ethics Board at Queen's University (DMED-1871-15).

Rubric development

In developing a rubric to evaluate the quality of EPAs, we started with the peer-reviewed literature to define the purpose for EPAs. The relevant literature[2,4,7,9,14,15] described the role of EPAs in CBME as operationalizing competencies to facilitate reliable assessment, which predicts future performance at a defined professional standard. Building on this, an effective EPA evaluation rubric would reliably measure the extent to which an EPA, as proposed, meets this purpose.

The elements of the published definition of EPAs were then reviewed to evaluate their contribution toward this purpose. In addition, given the purpose of the tool, six common misconceptions of EPAs described by ten Cate et al[4] were also reviewed for their role in evaluating EPA quality. D.T., A.T., R.E., and Y.S.P. all reviewed the elements of the definition and the six misconceptions identified by ten Cate et al and agreed that they aligned with the purpose and were each appropriate to be included in the rubric. Further, these constructs of EPAs organized into three categories: discrete units of work; entrustable and essential tasks of the profession; and education-focused considerations. To develop a criterion-based approach, descriptive anchors were created for each item to capture the degree of achievement across a five-point scale.

Expert consensus conference

A group of five health professions education scholars with experience in EPA development then applied the rubric to a panel of 31 internal medicine (IM) EPAs being developed for residency training in Canada. These EPAs had been selected collectively by clinical content experts to compose the full IM EPA panel and were presented to the education scholars for validation and quality review. When applying the rubric, the scholars had the full EPA report, including detailed descriptions and CanMEDS role mapping. Following application of the rubric, this group met to review the results and evaluate and revise the rubric. The rubric was reviewed for individual item wording, clarity and meaning of descriptive anchors, and overall structure. Revisions were deemed complete when all members of the panel agreed that the item captured the construct intended with clarity and accuracy.

Once consensus on the revisions was achieved, an Angoff approach was applied to set the cut score for each item. In this approach, the members of the education scholars team individually examined each item in the revised rubric and selected the score she/he felt represented the minimum score that would be considered consistent with the standard for EPAs. The group members then openly discussed their reasoning to address differences of opinion until unanimous consensus was achieved for each item. The consensus cut score for each item on the rubric therefore represented the minimum score for which the associated descriptive anchor would still be considered consistent with the defining qualities of EPAs. These item cut scores were averaged to determine an overall cut score for the rubric.

Training module development

To further standardize the use of the rubric, an online training module was developed to train and standardize users.[16] The module starts by reviewing the purpose and definition of EPAs. Subsequently, it reviews each item in the rubric in detail, clarifying the meaning and explaining the use of the descriptive anchors.

Rubric application

To evaluate the reliability of the rubric, the unrevised panel of EPAs initially presented to the education scholars was distributed to a team of four nonclinician support staff (education research assistants) and a team of four program directors or CBME leads for residency training (PDs) at Queen's University. Each participant completed the online training module and then applied the rubric to the full panel of EPAs.

Data analysis

To determine the reliability of evaluating EPA quality using the rubric and the associated sources of variance, we performed a generalizability study (G-study), where EPAs (e) were treated as the object of measurement, and the raters (r) and rubric items (i) were treated as facets. Raters were assumed to be random samples; items were assumed fixed. The G-study used a fully-crossed design, $e \times r \times i$.

The results for each of the two groups were analyzed to determine intraclass correlation coefficients (ICCs) for each item combination and for the rubric as a whole. ICC calculations were made using a two-way mixed-effects model, reporting consistency values for average measures. ICCs were calculated using SPSS 24 (IBM Corp., Armonk, New York). The item discrimination index was then calculated for the PDs using Stata 14 (Stata Corp., College Station, Texas).

To determine how closely ratings by nonclinician support staff correlated to ratings by PDs, correlation coefficients between the groups were measured. First, the average score for each item–EPA combination was determined for each group. The Pearson correlation coefficient was then determined for the groups for overall performance on the rubric and at the item level.

Comparison with revision decisions in actual EPA development

To evaluate the validity of the results from the PDs and nonclinician support

staff for making revision decisions in EPA development, their scores were compared with decisions made independently by the five education scholars involved in the actual EPA development process. For each group, the scores were averaged for each EPA; the average EPA scores were then compared with the overall cut score for the rubric to identify those in need of major revisions. These were then compared with the actual decisions made by the scholars regarding the need for major revision. Kappa was calculated to measure interrater agreement between the scholars and each of the PDs and nonclinician support staff using SPSS 24 (IBM Corp., Armonk, New York).

Results

PDs collectively rated EPAs with an average score of 4.22 (SEM ± 0.022) on the 5-point scale; the average score for individual raters ranged from 4.13 to 4.40. Their data had a median score of 4 with a mode of 5 and a scoring range from 1 to 5. Nonclinician support staff collectively rated the EPAs with an average score of 4.13 (SEM ± 0.022) with a range for individual raters of 3.79 to 4.42. Similarly, nonclinician support staff had a median of 4, a mode of 5, and a scoring range from 1 to 5.

Generalizability analysis

The generalizability analysis of the rubric demonstrated moderately good reliability amongst the nonclinician support staff with a phi-coefficient of 0.672 and excellent reliability amongst the PDs with a phi-coefficient of 0.837 (see Table 1).

Rater variance from nonclinician support staff was 9.10%, compared with the smaller 0.05% variance for the PDs, indicating greater variability in ratings among nonclinician support staff. (Some nonclinician support staff were more severe than others.)

The PD analysis also produced variance components attributable to interaction effects between EPAs and raters (9.80%), EPAs and items (11.10%), and raters and items (10.00%). The residual error in the analysis, the interaction effect between all three, produced a percent variance component of 40.00%, representing the largest source of variance in the analysis (see Table 1).

The generalizability analysis for the PDs was repeated excluding the data from three items on the rubric with poor interrater reliability. This subgroup analysis produced variance components for EPAs (24.30%), raters (0.00%), and items (18.6%). The analysis also showed a reduced interaction effect between raters and items with a percent variance component of 2.70%.

Item discrimination and ICCs

The average item discrimination index across all items for the PDs was 0.52 (SD ± 0.20) with a range of 0.11 to 0.73. Two items had discrimination indices below 0.30: Item 6 had an index of 0.18, and item 14 had an index of 0.11.

The overall ICCs for the PDs and the nonclinician support staff, and the ICCs for each item, are shown in Table 2. These results demonstrate very good reliability overall. At the item level, low reliability was seen for items 2 and 14 for the nonclinician support staff and for items 6 and 14 amongst the PDs. These items aligned with those items demonstrating poor item discrimination.

Correlations

The average scores for each item–EPA combination between the nonclinician support staff and the PDs were moderately to strongly correlated with a correlation coefficient of 0.602. At the item level, correlation of PD and RA scores did not reach statistical significance for 5 of the 14 items (see Table 3).

Comparing the average PD scores for each EPA with the decision for major revisions by the scholars' panel reveals excellent interrater reliability with a kappa of 0.852 (see Table 4). In contrast, the average nonclinician support staff scores poorly correlated with decisions for major revisions (kappa 0.384).

Discussion

In this study, we present a novel rubric designed for use in evaluating the alignment of EPAs with the elements of their definition as described in the peer-reviewed literature. The development and revision of the rubric was guided by international leaders in CBME and subjected to rigorous statistical analysis. The rubric provides criterion-based evaluation of EPAs and includes rater training. Further, the results obtained from the participants were compared with the revision decisions made on the EPAs in the actual development process. This comparison showed that the PDs' scores had excellent consistency (kappa) with the education scholars' decisions in making major revisions.

G-study analysis

This study provides evidence that the EQual rubric reliably measures EPA quality, as defined by the degree of alignment EPAs have with EPA standards described in the peer-reviewed literature. The generalizability theory analysis showed excellent reliability when four postgraduate PDs applied the rubric to stage-specific EPAs, with very little variance attributable to the raters themselves. Although excellent reliability

Table 1
Generalizability Study: Variance Components and Reliability

		Program directors		Nonclinician support staff	
Effects[a]	*df*	Variance component (SE)	% VC	Variance component (SE)	% VC
EPA (*e*)	30	0.150 (0.046)	17.00%	0.111 (0.038)	12.30%
Rater (*r*)	3	0.004 (0.009)	0.50%	0.082 (0.057)	9.10%
Item (*i*)	13	0.104 (0.048)	11.70%	0.107 (0.044)	11.80%
e × *r*	90	0.087 (0.017)	9.80%	0.114 (0.021)	12.60%
e × *i*	390	0.098 (0.014)	11.10%	0.068 (0.012)	7.60%
r × *i*	39	0.088 (0.022)	10.00%	0.036 (0.011)	4.00%
e × *r* × *i*, error	1,170	0.354 (0.015)	40.00%	0.384 (0.016)	42.50%
Phi-coefficient		0.837		0.672	

Abbreviations: SE indicates standard error; % VC, variance component expressed as a percentage of the total variance; EPA, entrustable professional activity.
[a]Full cross-generalizability study design used, with EPA crossed with raters and items.

Table 2
Interrater Reliability: Intraclass Correlation (ICC) of Ratings by Rubric Item

	ICC for program directors				ICC for nonclinician support staff			
Item(s)	N	ICC	95% confidence interval	*P* value	N	ICC	95% confidence interval	*P* value
1	31	0.729	(0.530 to 0.857)	< .001	31	0.612	(0.329 to 0.795)	< .001
2	31	0.399	(–0.041 to 0.682)	.035	31	–0.041	(–0.803 to 0.96)	.534
3	31	0.675	(0.437 to 0.828)	< .001	31	0.546	(0.213 to 0.760)	.002
4	31	0.853	(0.746 to 0.922)	< .001	31	0.502	(0.138 to 0.737)	.006
5	31	0.805	(0.663 to 0.897)	< .001	31	0.666	(0.422 to 0.824)	< .001
6	31	0.142	(–0.486 to 0.546)	.286	31	0.422	(–0.001 to 0.694)	.025
7	31	0.622	(0.346 to 0.800)	< .001	31	0.707	(0.493 to 0.845)	< .001
8	31	0.800	(0.654 to 0.894)	< .001	31	0.635	(0.367 to 0.807)	< .001
9	31	0.768	(0.598 to 0.877)	< .001	31	0.775	(0.611 to 0.881)	< .001
10	31	0.829	(0.703 to 0.909)	< .001	31	0.782	(0.623 to 0.885)	< .001
11	31	0.711	(0.500 to 0.847)	< .001	31	0.788	(0.633 to 0.888)	< .001
12	31	0.467	(0.025 to 0.702)	.020	31	0.548	(0.217 to 0.761)	.002
13	31	0.518	(0.166 to 0.745)	.004	31	0.412	(–0.019 to 0.689)	.029
14	31	–0.152	(–0.995 to –0.391)	.662	31	–0.081	(–0.872 to 0.428)	.583
Overall	434	0.723	(0.678 to 0.764)	< .001	434	0.676	(0.623 to 0.723)	< .001

was achieved, the variance component attributable to EPAs, the object of measure, was only 17%. The variance due to the interaction between the rubric items and raters was 10%.

Table 3
Correlation Coefficients for Scores From PDs and Nonclinician Support Staff by Item

Item	Correlation coefficient[a]	*P* value
1	0.49	.005
2	0.06	.757
3	0.16	.407
4	0.69	< .001
5	0.70	< .001
6	0.28	.121
7	0.54	.002
8	0.71	< .001
9	0.72	< .001
10	0.81	< .001
11	0.77	< .001
12	0.33	.067
13	0.33	.069
14	0.45	.011
Overall	0.60	< .001

Abbreviation: PDs indicates program directors or competency-based medical education leads.
[a]Correlation coefficients relate PD scores to nonclinician support staff scores for each item across all EPAs.

To address this variance component, we examined the rubric for item discrimination index and ICC by item. Items 2, 6, and 14 on the rubric showed low interrater reliability. Further, items 6 and 14 showed very low item discrimination indices. Despite a well-defined role in the blueprint of the rubric, expert-guided item revision, and rater training, these items performed poorly from a psychometric standpoint. Further, these limitations were noted across the PDs and nonclinician support staff, as well as in the pilot with medical education scholars (data not shown).

Reanalysis of our results excluding the data from these three rubric items reduced the variance component due to the rater–item interaction and increased the variance attributable to the EPA. Although this analysis was completed post hoc on a subset of data, it suggests that eliminating these three items would improve the psychometric characteristics of the rubric (internal structure validity evidence).

Items not meeting the grade

Item 2, "independently executable to achieve a defined clinical outcome," asks raters to assess to what degree an activity is a stand-alone task in practice for achieving a particular outcome not explicitly defined in the EPA. Agreement amongst raters requires a shared understanding of the clinical outcome to be achieved. For the example of performing a procedure, is the outcome simply technical completion or is it a therapeutic or diagnostic outcome? Further, agreement demands that the raters make a judgment about the degree to which the outcome is achieved by the task alone. In the procedural EPA example, the technical performance is discrete and stand-alone, but it occurs in the context of caring for a patient, often one who is acutely ill. Raters would rate this item differently depending on whether they view the procedure in isolation or as part of a complex and integrated patient care plan. Although independently executable is clearly a key element of EPAs from an assessment standpoint, it is one which, by nature, eludes quantitative description.

Item 6 asks whether the EPA is distinguishable from the other EPAs in the framework. Comparing each EPA with the others is tedious work that likely was approached differently by raters. Additionally, in this study we examined a panel of stage-specific EPAs, which include EPAs assigned to intermediate transition points in residency training in addition to the transition from residency to practice. A complete set of stage-specific EPAs is likely to include nested EPAs at early stages that will share overlap with EPAs from the advanced stages of training. This raises the question of whether the challenges with this item were related to the subject matter (which would be captured in $e \times i$ and $r \times e$ interactions' attributable variance).

Item 14 was the most surprising in that the task assigned, "identify the presence of adjectives or adverbs in the EPA that describe proficiency," appears a simple and concrete task. As with items 2 and 6, this challenge persisted across the pilot study, and the analyses with the PDs and the nonclinician support staff. The poor agreement may have represented difficulty distinguishing adjectives that characterize the activity, and adjectives or adverbs that characterize the learner's performance. In truth, in writing EPAs, adverbs, more often than adjectives, are the qualifiers that refer to proficiency.

All three of these items address important aspects of EPAs for those involved in EPA development and evaluation. Our study

Table 4

Comparison With Revision Decisions in Actual EPA Development[a,b]

EPA number	PD score[c]	PD score below cut score of 4.07	Required major revision as determined by scholars[d]
1	4.54	No	No
2	4.52	No	No
3	4.63	No	No
4	4.38	No	No
5	4.57	No	No
6	4.61	No	No
7	4.32	No	No
8	4.57	No	No
9	3.89	Yes	Yes
10	3.30	Yes	Yes
11	4.61	No	No
12	4.59	No	No
13	4.63	No	No
14	4.46	No	No
15	3.70	Yes	Yes
16	4.50	No	No
17	4.68	No	No
18	4.54	No	No
19	**3.63**	**Yes**	**No**
20	4.18	No	No
21	3.59	Yes	Yes
22	3.55	Yes	Yes
23	3.46	Yes	Yes
24	3.66	Yes	Yes
25	3.77	Yes	Yes
26	**4.61**	**No**	**Yes**
27	4.36	No	No
28	4.50	No	No
29	4.39	No	No
30	3.93	Yes	Yes
31	4.29	No	No

Abbreviations: EPA indicates entrustable professional activity; PD, program director or competency-based medical education lead.

[a]Shading denotes recommendation by PDs to revise.

[b]Outlining denotes discordance between PDs and revision decisions.

[c]PD scores generated using revised, 14-item rubric.

[d]Requirement for revision was determined by the education scholars at the consensus conference, independent of subsequent PD and nonclinician support staff evaluations.

suggests that these constructs are resistant to reliable quantification and are likely better evaluated by a different approach.

Identifying EPAs requiring revision

Our study also showed that the EQual rubric is effective in identifying EPAs needing major revision. Applying the overall cut score set for the rubric, the ratings from the PDs independently identified 9 of the 10 EPAs in the framework that had required major revision, as determined by the education scholars panel. Of the 21 EPAs that required minor or no revision, 20 were rated by the PDs to have met the cut score. This is an important finding for those involved in EPA development and evaluation as it demonstrates the ability to selectively identify those EPAs requiring more attention.

Extending the use of the rubric to nonclinician support staff

The reasonable reliability achieved by the nonclinician support staff suggests that the rubric could be used by them in support of EPA evaluation and in research, with some limitations. Although the nonclinician support staff achieved a correlation coefficient of 0.602 overall with the PDs, at the item level we found multiple items with poor correlations, not achieving statistical significance (see Table 3). These findings indicate that education-trained nonclinician support staff using the rubric can provide some global information about a panel of EPAs, but their results should not be used to selectively identify EPAs for revision, or to guide aspects of the EPAs to revise. This area requires further investigation.

Implications

Our study describes a reliable tool to evaluate EPA quality using a criterion-based approach grounded in the peer-reviewed literature. The EQual rubric provides those developing EPAs, whether at the local, regional, or national level, a tool to help validate the EPAs being proposed and a guide for directing revisions. From a program evaluation standpoint, EQual offers groups revising EPA-based curricula a tool to better understand shortcomings and strengths in educational frameworks. Finally, for education researchers, the rubric provides a tool to aid with validation studies in EPA development work, and with studies examining the value of extending EPA use beyond the clinical environments for which they were intended.

Efforts are under way to refine this work and address some areas for improvement, as part of a future study. First, the rubric was used for one set of EPAs, in one specialty. Additionally, the EPAs evaluated in this study were stage-specific EPAs as defined by the Royal College of Physicians and Surgeons of Canada. Although stage-specific EPAs align well with the literature definition of EPAs, they differ from typical end-of-training EPAs by including intermediate EPAs that may be nested within larger, end-of-training EPAs. Particularly considering item 6 in the rubric, this could potentially influence results. This would potentially limit the generalizability of these results to other training jurisdictions. Finally, the revised rubric in which items 2, 6, and 14 were eliminated was analyzed post hoc. These results need to be prospectively validated to demonstrate that the shortened rubric indeed provides reliable and useful results.

Conclusions

Our study provides solid evidence supporting the use of the EQual rubric for reliably measuring EPA quality, as defined in the literature. Its performance characteristics suggest it is an excellent tool for users developing and evaluating EPAs; its application helps identify EPAs in need of major revision when applied by clinician–educators to a panel of EPAs. It also provides information to EPA developers regarding areas of strengths and weaknesses in the structure and content of EPAs. It is the first EPA evaluation tool to use descriptive anchors in its rating scales and provide online rater training for its use. However, its use by nonclinician support staff needs additional exploration. Finally, as a research tool, our findings indicate that EQual has potential for use in validating methods used for EPA development.

Acknowledgments: The authors would like to thank Damon Dagnone, Peggy De Jong, Janet Lui, and Mark Ropeleski for their time and support in developing and testing the rubric.

Funding/Support: The project was funded by the Southeastern Ontario Academic Medical Organization (SEAMO) Endowed Scholarship and Education Fund.

Other disclosures: None reported.

Ethical approval: This study was approved by the Queen's University Health Science Research Ethics Board, study code DMED-1871-15.

D.R. Taylor is associate professor, Department of Medicine, Queen's University School of Medicine, Kingston, Ontario, Canada.

Y.S. Park is assistant professor, Department of Medical Education, University of Illinois at Chicago, Chicago, Illinois.

R. Egan is assistant professor and director, Office of Health Science Education, Queen's University School of Medicine, Kingston, Ontario, Canada.

M.-K. Chan is associate professor, Department of Pediatrics and Child Health, University of Manitoba, Winnipeg, Manitoba, Canada, and clinician educator, Royal College of Physicians and Surgeons of Canada, Ottawa, Ontario, Canada.

J. Karpinski is associate professor, Department of Medicine, University of Ottawa, Ottawa, Ontario, Canada, and associate director, Specialties Unit, Royal College of Physicians and Surgeons of Canada, Ottawa, Ontario, Canada.

C. Touchie is associate professor, Department of Medicine, University of Ottawa, Ottawa, Ontario, Canada, and chief medical education advisor, Medical Council of Canada, Ottawa, Ontario, Canada.

L.S. Snell is professor, Department of Medicine, McGill University, Montreal, Quebec, Canada, and senior clinician educator, Royal College of Physicians and Surgeons of Canada, Ottawa, Ontario, Canada.

A. Tekian is professor, Department of Medical Education, University of Illinois at Chicago, Chicago, Illinois.

References

1 Cate ten O, Snell L, Carraccio C. Medical competence: The interplay between individual ability and the health care environment. Med Teach. 2010;32(8):669–675.
2 Ten Cate O, Billett S. Competency-based medical education: Origins, perspectives and potentialities. Med Educ. 2014;48:325–332.
3 Association of American Medical Colleges. Core entrustable professional activities for entering residency: Curriculum developer's guide. Washington, DC: Association of American Medical Colleges; 2014. https://members.aamc.org/eweb/upload/Core%20EPA%20Curriculum%20Dev%20Guide.pdf. Accessed July 19, 2017.
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;37:983–1002.
5 Caverzagie KJ, Cooney TG, Hemmer PA, Berkowitz L. The development of entrustable professional activities for internal medicine residency training: A report from the Education Redesign Committee of the Alliance for Academic Internal Medicine. Acad Med. 2015;90:479–484.
6 Cate ten O, Hart D, Ankel F, et al. Entrustment decision making in clinical training. Acad Med. 2016;91(2):191–198.
7 Cate ten O. Entrustability of professional activities and competency-based training. Med Educ. 2005;39(12):1176–1177.
8 Touchie C, ten Cate O. The promise, perils, problems and progress of competency-based medical education. Med Educ. 2016;50:93–100.
9 ten Cate O, Scheele F. Competency-based postgraduate training: Can we bridge the gap between theory and clinical practice? Acad Med. 2007;82:542–547.
10 Tekian A. Are all EPAs really EPAs? Med Teach. 2017;39:232–233.
11 Touchie C, Boucher A, Allen T, et al. Entrustable Professional Activities for the Transition From Medical School to Residency. Ottawa, Ontario, Canada: Association of Faculties of Medicine of Canada; 2016. https://afmc.ca/sites/default/files/documents/AFMC_Entrustable_Professional_Activities_EN_0.pdf. Accessed July 19, 2017.
12 Wisman-Zwarter N, van der Schaaf M, Ten Cate O, Jonker G, van Klei WA, Hoff RG. Transforming the learning outcomes of anaesthesiology training into entrustable professional activities: A Delphi study. Eur J Anaesthesiol. 2016;33:559–567.
13 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;31:518–523.
14 Frank JR, Snell LS, Cate OT, et al. Competency-based medical education: Theory to practice. Med Teach. 2010;32:638–645.
15 Holmboe ES, Sherbino J, Long DM, Swing SR, Frank JR. The role of assessment in competency-based medical education. Med Teach. 2010;32:676–682.
16 Taylor DR. EQal rubric training video for EPA evaluation. Published May 30, 2017. https://youtu.be/yQZuWdzkQKM. Accessed August 8, 2017.

Appendix 1

EQual Rubric Items[a]

Discrete Activity

1. This EPA has a clearly defined beginning and end[4]
2. This EPA is independently executable to achieve a defined clinical outcome[7]
3. This EPA is specific and focused[4]
4. This EPA is observable in process[7,9]
5. This EPA is measureable in outcome[7,9]
6. This EPA is clearly distinguished from other EPAs in the framework[4]

Entrustable, Essential, and Important Task of the Profession

7. This EPA describes work that is essential and important to the profession[7,9]
8. Performing this EPA leads to recognized output or outcome of labor[7,9]
9. The performance of this EPA in clinical practice is restricted to qualified personnel[7,9]
10. This EPA addresses professional work that is suitable for entrustment[4]

EPA as Educational Tool

11. This EPA requires the application of knowledge, skills, and/or attitudes (KSAs) acquired through training[7,9]
12. This EPA involves application and integration of multiple domains of competence[7,9]
13. The EPA title describes a task, not qualities or competencies of a learner[4]
14. This EPA describes a task and avoids adjectives (or adverbs) that refer to proficiency[4]

[a]For complete rubric including descriptive anchors see Appendix 2.

Appendix 2

Entrustable Professional Activities (EPAs) as Discrete Units of Work

This EPA has a clearly defined beginning and end

Neither the beginning nor the end of the activity is clearly defined		The beginning OR the end is clearly defined but not both		The beginning and end are both clearly defined
1	2	3	4	5

This EPA is independently executable to achieve a defined clinical outcome

Routinely depends on multiple other contributing tasks/activities	Routinely depends on one other contributing task/activity	Can be independent, but commonly depends on other tasks/activities to achieve its clinical outcome	Typically independent, but infrequently depends on other tasks/activities to achieve its clinical outcome	Independent of other tasks/activities to achieve its clinical outcome
1	2	3	4	5

This EPA is specific and focused

Describes a large, general area of practice or describes domains of competence	Is a general category of work that serves a broad purpose	Is a general category of work that serves a clear and focused purpose	Includes a few closely-related units of work that serve a common, clear and focused purpose	Is specific work that serves a clear and focused purpose
1	2	3	4	5

This EPA is observable in process

The activity cannot be observed or monitored	Parts of the activity can be monitored, but only indirectly	Some parts of the activity can be directly observed	Most of the activity can be directly observed, but not the entire activity	The activity can be observed in all aspects from beginning to end
1	2	3	4	5

This EPA is measureable in outcome

The outcome of the work cannot be described or measured	Limited aspects of the outcome can be inferred from indirect assessment but not direct measurement	The outcome of the work can be inferred, but not directly described or measured	The outcome of the work can be largely described and/or measured directly	The outcome of the work can be fully described and/or measured directly
1	2	3	4	5

(*Appendix continues*)

Appendix 2, (Continued)

This EPA is clearly distinguished from other EPAs in the framework

Cannot be meaningfully distinguished from one or more of the other EPAs	Has clear similarity or overlap with one or more of the other EPAs	Has similarity with other EPAs in the framework, but there are also some clear distinguishing features	Has some similarity with one or more EPAs in the framework, but there are clear and important distinguishing features	Has no apparent overlap with other EPAs in the framework
1	2	3	4	5

EPAs as Entrustable, Essential, and Important Tasks of the Profession

This EPA describes work that is essential and important to the profession

Very low importance to professional practice	Limited importance; is non-essential to practice	Important but professional practice could succeed without it	Important and is expected for professional practice	Very important and essential to professional practice
1	2	3	4	5

Performing this EPA leads to recognized output or outcome of labour

No discernable product or recognized outcome from the work	Variably produced outcome but it is not clearly attributable to the work	Variably produced outcome attributable to the work, or a typically produced outcome not clearly attributable to the work	Typically a defined outcome attributable to the work	A clear and defined outcome consistently produced from the work
1	2	3	4	5

The performance of this EPA in clinical practice is restricted to qualified personnel

Is routinely done by untrained persons	Requires limited training to perform	Requires training to perform	Requires training and qualification/certification to perform	Exclusively performed by trained and qualified individuals within the profession
1	2	3	4	5

This EPA addresses professional work that is suitable for entrustment

Has no influence on the well-being of patients or the public as a whole	May contribute to health care system, but only minimally or indirectly influences the care of patients or the public as a whole	Contributes to the well-being of the public as a whole, but lacks direct influence on clinical care	Expected of a physician and contributes to safe clinical practice, but is not clinical care itself	Clearly expected of a physician as part of delivering competent clinical care
1	2	3	4	5

EPAs' Curricular Role

This EPA requires the application of knowledge, skills, and/or attitudes (KSAs) acquired through training

The KSAs required for the task are not acquired through training	Training adds somewhat to the KSAs required for the task	The KSAs required for the task require training, but success is strongly influenced by non-trainable qualities	The task is largely dependent on trainable KSAs for success and is influenced only modestly by non-trainable physician qualities	The task is completely dependent on KSAs acquired through training for success
1	2	3	4	5

This EPA involves application and integration of multiple domains of competence

Does not reflect any identified domain of competence	Reflects only one domain of competence	Reflects one domain of competence, although other domains may be identifiable	Requires integration of multiple domains to perform but with one domain dominant	Involves the integration of multiple domains of competence
1	2	3	4	5

The EPA title describes a task, not qualities or competencies of a learner

Describes a single quality/ competency of a clinician without describing application in clinical work	Describes a single quality/ competency of a clinician and references application in a clinical context	Describes the clinician who integrates multiple qualities/competencies but does not describe clinical application	Describes the clinician who integrates multiple competencies and also describes the associated professional activity	Describes only a professional activity
1	2	3	4	5

This EPA describes a task and avoids adjectives (or adverbs) that refer to proficiency

Employs adjectives that focus it primarily on proficiency and not the task	Employs adjectives that focus it primarily on proficiency, but also describes the task	Employs adjectives referencing proficiency, but overall primarily describes the task	Employs no adjectives referencing proficiency; it does imply some aspects of proficiency of the learner	Employs no adjectives referencing proficiency and does not imply proficiency of the learner performing the task
1	2	3	4	5