



Rating the Quality of Entrustable Professional Activities: Content Validation and Associations with the Clinical Context

Jason A. Post, M.D.¹, Christopher M. Wittich, M.D., Pharm.D.¹, Kris G. Thomas, M.D.¹, Denise M. Dupras, M.D., Ph.D.¹, Andrew J. Halvorsen, M.S.¹, Jay N. Mandrekar, Ph.D.², Amy S. Oxentenko, M.D.¹, and Thomas J. Beckman, M.D.¹

¹College of Medicine, Department of Internal Medicine, Mayo Clinic, Rochester, MN, USA; ²College of Medicine, Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA.

BACKGROUND: Entrustable professional activities (EPAs) have been developed to assess resident physicians with respect to Accreditation Council for Graduate Medical Education (ACGME) competencies and milestones. Although the feasibility of using EPAs has been reported, we are unaware of previous validation studies on EPAs and potential associations between EPA quality scores and characteristics of educational programs.

OBJECTIVES: Our aim was to validate an instrument for assessing the quality of EPAs for assessment of internal medicine residents, and to examine associations between EPA quality scores and features of rotations.

DESIGN: This was a prospective content validation study to design an instrument to measure the quality of EPAs that were written for assessing internal medicine residents.

PARTICIPANTS: Residency leadership at Mayo Clinic, Rochester participated in this study. This included the Program Director, Associate program directors and individual rotation directors.

INTERVENTIONS: The authors reviewed salient literature. Items were developed to reflect domains of EPAs useful for assessment. The instrument underwent further testing and refinement. Each participating rotation director created EPAs that they felt would be meaningful to assess learner performance in their area. These 229 EPAs were then assessed with the QUEPA instrument to rate the quality of each EPA.

MAIN MEASURES: Performance characteristics of the QUEPA are reported. Quality ratings of EPAs were compared to the primary ACGME competency, inpatient versus outpatient setting and specialty type.

KEY RESULTS: QUEPA tool scores demonstrated excellent reliability (ICC range 0.72 to 0.94). Higher ratings were given to inpatient versus outpatient (3.88, 3.66; $p=0.03$) focused EPAs. Medical knowledge EPAs scored significantly lower than EPAs assessing other competencies (3.34, 4.00; $p<0.0001$).

CONCLUSIONS: The QUEPA tool is supported by good validity evidence and may help in rating the quality of EPAs developed by individual programs. Programs should take care when writing EPAs for the outpatient setting or to assess medical knowledge, as these tended to be rated lower.

KEY WORDS: medical education—assessment/evaluation; medical education—graduate; medical education—quantitative methods.

J Gen Intern Med 31(5):518–23

DOI: 10.1007/s11606-016-3611-8

© Society of General Internal Medicine 2016

INTRODUCTION

The Accreditation Council for Graduate Medical Education (ACGME) outcomes project marked the transition from process-based to outcomes-based assessment in medical education.¹ This project introduced the six core competencies of patient care, medical knowledge, systems-based practice (SBP), practice-based learning and improvement (PBLI), professionalism, and interpersonal and communication skills (ICS). While these competencies increased emphasis on educational outcomes, assessing individual learner performance in the competencies can be difficult.²

With the Next Accreditation System (NAS), the ACGME and the American Board of Medical Specialties introduced 22 competency-based developmental outcomes, referred to as sub-competencies or reporting milestones.³ Advantages of sub-competencies include the incorporation of specific behavioral anchors and the ability to gauge learners' progression over time. However, utilizing sub-competencies for clinical assessments may be challenging as they were designed primarily as a tool for reporting, not assessment.⁴

Consequently, entrustable professional activities (EPAs) and observable practice activities (OPAs) have been developed to operationalize the assessment of sub-competencies in the clinical setting.⁵ EPAs incorporate entrustment decisions—which are determinations regarding levels of responsibility—into the assessment of clinical performance.^{6–9} Ultimately, EPAs are useful for clinical assessment because they incorporate multiple competencies, observable behaviors, and the ability to draw conclusions about

This work was presented at the Spring Alliance for Academic Internal Medicine Meeting, April 2015.

Electronic supplementary material The online version of this article (doi:10.1007/s11606-016-3611-8) contains supplementary material, which is available to authorized users.

Received June 29, 2015

Revised November 18, 2015

Accepted January 27, 2016

Published online February 22, 2016

accountability and trustworthiness. OPAs differ from EPAs in that they are more context or content specific than EPAs, but are used in a similar manner to EPAs for learner assessment.¹⁰ Observable practice activities were originally described as a way to assess learners' observed performance, which in turn serves to inform entrustment within EPAs.¹⁰ However, EPAs expand beyond OPAs, by addressing activities that society trusts all physicians to do,⁵ and by allowing assessment of these activities over the course of a resident's training.

Although experts have clearly articulated the importance of EPAs in outcomes-based assessment, there remains a need for quantitative research on developing and assessing the quality of EPAs.¹¹ Literature regarding EPAs is just emerging, with recent studies exploring the feasibility and process of using EPAs for assessment.^{12–15}

Several national organizations have published EPAs for assessing medical students or internal medicine residents.^{16–18} While these EPAs are informative and useful, it is recognized that individual programs will need to develop additional EPAs at local levels in order to determine learners' progress and readiness for advancement in their unique clinical environments.¹⁷ Consequently, there is a need for validated tools that will help educational programs create high-quality EPAs. Our objectives were to: 1) develop and validate an instrument for assessing the quality of EPAs for milestones-based assessment, 2) describe the features of high quality EPAs, and 3) examine associations between EPA quality scores and the characteristics of rotations from which they originated, in order to determine whether certain learning environments were more conducive to writing EPAs for use in resident assessment.

METHODS

Setting and Participants. This was a prospective content validation study to identify and rate the quality of locally developed EPAs that were designed to assess the competence of internal medicine residents at our institution in 2014. Our program consists of 168 resident physicians, 33 rotation directors and ten associate program directors (APDs). All rotation directors were provided with basic education regarding the definition and purpose of EPAs at a noon conference. This one-hour educational session involved an interactive presentation by our local expert in EPAs, who reviewed the literature regarding EPAs and discussed the purpose of creating EPAs for evaluation. This education was reinforced with a subsequent e-mail message. Rotation directors were then instructed to write and submit EPAs that pertained to their subspecialty rotations, resulting in 229 locally developed EPAs (see [electronic addendum](#)). We rated the quality of these EPAs as well as previously published EPAs¹⁷ with an instrument we developed for the purpose of this study, which was

deemed exempt by the Mayo Clinic Institutional Review Board.

Instrument Development. To develop an instrument for rating the quality of EPAs, we created a team of raters separate from the rotation directors who had submitted the EPAs. This team of raters consisted of five APDs (authors JP, CW, DD, KT, and TB) who had previous experience in graduate medical education, resident assessment, scale development and validation.

Identifying Item Content. The team reviewed salient literature on the topic of EPAs,^{2,5–9,11,13–15,19,20} and then met to discuss the qualities of EPAs that would be useful in assessment. After repeated review and discussions, the team reached consensus on the following essential domains of an EPA: 1) Focus, 2) Observable, 3) Clear Intention, 4) Realistic, 5) Articulates Trustworthiness, 6) Generalizable Across Rotations, and 7) Integrates Multiple Competencies. For each domain, every team member proposed three potential items, resulting in 105 candidate items equally distributed across these seven domains. Subsequently, the team reached agreement on three items for each EPA quality domain. Items were structured on a five-point scale (1 = strongly disagree, 5 = strongly agree).

Determining Item Reliability. To pilot test the instrument, all team members rated a convenience sample of ten locally developed EPAs. Intraclass correlation coefficients (ICCs) were calculated to determine inter-rater reliability for each item, with ICCs < 0.4 considered to be poor, 0.4 to 0.75 considered to be fair to good, and > 0.75 considered to be excellent.²¹ This first round of pilot testing revealed good reliability (ICC range 0.72 to 0.94) for the domains of Focus, Observable, Realistic, Generalizable, and Integrates Multiple Competencies. However, poor reliability (ICC range 0.24 to 0.61) existed for the domains of Clear Intention and Articulates Trustworthiness. Despite additional pilot testing by rating ten different EPAs, the domains of Clear Intention and Articulates Trustworthiness continued to perform poorly, so they were dropped from the instrument, leaving the five domains of Focus, Observable, Realistic, Generalizable, and Integrates Multiple Competencies.

Rating the Quality of EPAs. The final instrument was called the Quality of EPA (QUEPA). Since the items in these five domains had demonstrated excellent inter-rater reliability, the team members were randomly assigned and independently rated equally sized subsets ($n=46$) of all locally developed EPAs using the QUEPA tool. Each team member also rated the AAIM End of Training EPAs (EOTEPA) to assess the application of QUEPA to EPAs that were not locally developed. Study data were collected and managed using REDCap

electronic data capture tools hosted at Mayo Clinic Rochester.²² REDCap (Research Electronic Data Capture) is a secure, web-based application designed to support data capture for research studies.

Study Variables. Independent variables were primary ACGME competency (patient care, ICS, professionalism, medical knowledge, SBP, and PBLI), practice location (inpatient, outpatient), rotation type (general medicine, medicine subspecialty, non-medicine specialty), and activity type (EPA vs OPA). For primary ACGME competency and activity type, each team member classified every locally developed EPA. Although EPAs reflect multiple competencies, the group assessed which competency they felt was most represented by each EPA. This is a relevant step, as many programs will likely use performance in EPAs to assess resident progress in the sub-competencies, and high quality EPAs will need to be created to inform these decisions. In circumstances where the team did not agree, author JP adjudicated the final decision. The outcome variable that was utilized to calculate associations with the clinical context was the average QUEPA score (scale 1–5).

Data Analysis. Characteristics of rotation directors were described using descriptive statistics including sex, academic rank, time on faculty, and proceduralist versus non-proceduralist specialty. For the purposes of this study, non-proceduralist was defined as general medicine, allergy, endocrinology, hematology, infectious diseases, nephrology, neurology, preventive medicine, and rheumatology; and proceduralist was defined as cardiology, gastroenterology, emergency medicine, and pulmonary critical care. The dimensionality of the final 15-item QUEPA instrument was determined using factor analysis with orthogonal rotation. To account for the clustering of multiple ratings within raters and EPAs, we generated an adjusted correlation matrix using generalized estimating equations. This adjusted correlation matrix was then used to perform confirmatory factor analysis with orthogonal rotation. In addition, for a sensitivity analysis, we performed factor analysis using an unadjusted correlation matrix and within rater and EPA combinations separately. Factors were identified using the minimal proportion criteria. The threshold for item retention was factor loading > 0.4. We then calculated the percentage of shared variance that the extracted factors contributed to the original variables. Internal consistency reliability for items within factors and overall was calculated using Cronbach α , with coefficients > 0.7 considered to be acceptable. ANOVA models with random effects for rotation directors were used to compare overall QUEPA scores for subcategories within ACGME competency, practice location, practice type and activity type. The level for statistical significance was set at $\alpha=0.05$. Statistical analyses were conducted using SAS 9.3 (SAS Institute Inc., Cary, North Carolina).

RESULTS

Summary of Rotation Directors who Submitted EPAs. Of the 33 rotation directors, 15 (45%) provided a total of 229 EPAs for 20 rotations. Participating rotation directors were predominantly female ($N=9$, 60%) and from non-proceduralist specialties (10, 67%). Academic ranks of the rotation directors were: Instructor (2, 13%), Assistant Professor (11, 73%), and Associate Professor (2, 13%), with an average of 9.3 years on faculty ($SD=7.1$).

Factor Analysis and Internal Consistency Reliability. Factor analysis revealed the five essential domains of an EPA reduced to four distinct factors (number of items) when applying the QUEPA instrument: Realistic and Generalizable (6), Observable (3), Focused (3), and Multiple Competencies (3). These factors explained 100% of the shared variance among the original items. Internal consistency for the individual factors was excellent (Cronbach α range: 0.950 to 0.990) (Table 1).

Table 1. Quality of Entrustable Professional Activity (QUEPA) Rating Instrument: Item Loadings and Internal Consistency Reliability

Factors and items	Item loadings			
	Factor 1	Factor 2	Factor 3	Factor 4
Focused				
The EPA is focused	0.056	0.182	0.958	-0.088
The EPA is not too broad in scope	0.742	0.175	0.959	-0.853
Observable				
The EPA addresses a single activity	0.041	0.200	0.935	-0.960
The EPA addresses a visible action	0.142	0.926	0.225	0.138
The EPA can be seen	0.134	0.930	0.220	0.136
The EPA is observable	0.134	0.929	0.210	0.145
Realistic and Generalizable				
The EPA is a realistic activity for an internist	0.822	0.399	-0.087	0.003
The EPA deals with an activity pertinent to internal medicine	0.827	0.382	-0.073	0.002
The EPA is an activity that an internist would be expected to perform	0.832	0.389	-0.081	0.023
The EPA is generalizable to different aspects of medical practice	0.918	-0.058	0.144	0.112
The EPA transfers to multiple settings	0.917	-0.065	0.140	0.108
The EPA is useful across many specialties	0.920	-0.056	0.154	0.108
Multiple competencies				
The EPA incorporates multiple abilities	0.086	0.108	-0.093	0.973
The EPA reflects multiple competencies	0.055	0.108	-0.092	0.975
The EPA requires integration of knowledge, skills, and attitudes	0.106	0.147	-0.073	0.944
Cronbach alpha	0.950	0.997	0.990	0.989

Associations Between QUEPA Scores and Competency or Clinical Context. Statistically significant associations were seen between mean QUEPA score and ACGME competency ($p < 0.0001$), activity type ($p < 0.0001$), and practice location ($p = 0.03$). Across the six competencies, ICS EPAs were rated highest (4.00), while medical knowledge EPAs were rated lowest (3.34). Entrustable Professional Activities scored higher on average (4.00) than Observable Practice Activities (3.51), and EPAs from inpatient rotations scored higher on average (3.88) than their outpatient counterparts (3.66). There were no significant associations seen between QUEPA scores and rotation type (Table 2).

Ratings of End of Training EPAs. QUEPA ratings of EOTEPAs showed good to excellent inter-rater reliability (overall mean ICC=0.75) (Table 3). The overall mean QUEPA score for EOTEPAs was 3.83, with the highest rated factor being Multiple Competencies (4.53), and the lowest rated factor being Focused (2.60). The highest rated individual EOTEPA was “facilitate family meetings” (4.43), and the lowest rated EOTEPA was “Improve the quality of health care at both the individual and systems level” (3.31).

DISCUSSION

To our knowledge, this is the first study to create and validate an instrument for rating the quality of EPAs. This instrument also allowed us to determine associations between EPA quality and characteristics of the clinical context in an internal medicine residency training program. We found that EPAs written to assess medical knowledge scored lowest among the six ACGME competencies; that EPAs for outpatient rotations scored lower than EPAs for inpatient rotations; and that EPAs

were rated more highly than OPAs. These findings have implications for creating and refining EPAs for learner assessment in residency training.

We found a significant difference in QUEPA scores, with EPAs assessing medical knowledge scoring lower than the five other competencies. A potential reason for this finding is that EPAs are designed for evaluating activities, such as communication, that integrate multiple skills and tasks⁶. Previous studies have found that assessing the medical knowledge of trainees is challenging^{23–25}; therefore, it is understandable that faculty would have difficulty creating EPAs for assessing medical knowledge. Furthermore, it can be challenging in the clinical setting to distinguish resident performance in the medical knowledge competency versus the patient care competency based solely on observation. Hence, the most effectively written EPAs are more likely to describe an observable patient care behavior as opposed to the simple recollection of facts.

We also found that EPAs written for inpatient rotations were rated higher than EPAs written for outpatient rotations. This finding supports a previous study showing that EPAs written for the inpatient setting had the highest ratings.¹³ One explanation for this finding is that learner–patient interactions are more likely to be observed in the inpatient setting than in the outpatient setting. Traditionally, hospital rounds create opportunities to see a learner’s performance at the bedside, whereas outpatient clinics often require only the confirmation of key findings by the faculty, which may allow for less frequent, direct observation. Moreover, senior residents usually function more independently in the clinic and are thus supervised less directly by faculty members. Consequently, we suspect that faculty members who teach primarily in the hospital were better poised to write higher quality EPAs that focused on observed behaviors. This finding suggests that clinic-based faculty might consider models of teaching that include more direct observation.

Table 2. Associations Between Quality of Entrustable Professional Activity (QUEPA) Scores with ACGME Competency, Rotation Practice Location, Rotation Specialty, and Entrustable Professional Activity (EPA) Versus Observable Professional Activity (OPA)

Characteristic	Number (%) of submitted EPAs and OPAs	QUEPA Mean (SD)	P value
ACGME competency			< 0.0001
Patient care	107 (46.7)	3.71 (0.54)	
Medical knowledge	19 (8.3)	3.34 (0.72)	
Interpersonal and communication skills	34 (14.8)	4.00 (0.45)	
Professionalism	17 (7.4)	3.93 (0.68)	
Systems-based practice	39 (17.0)	3.83 (0.55)	
Practice-based learning and improvement	13 (5.7)	3.95 (0.45)	
Rotation practice location			0.03
Inpatient	114 (49.8)	3.88 (0.55)	
Outpatient	115 (50.2)	3.66 (0.57)	
Rotation specialty type			0.72
General medicine	44 (19.2)	3.84 (0.46)	
Medicine sub-specialty	156 (68.1)	3.75 (0.59)	
Non-medicine specialty	29 (12.7)	3.81 (0.62)	
Activity type			< 0.0001
Entrustable Professional Activity (EPA)	123 (53.7)	4.00 (0.47)	
Observable Practice Activity (OPA)	106 (46.3)	3.51 (0.57)	

Table 3. End of Training EPA Ratings Using QUEPA

Factors and items	QUEPA Mean Scores (SD)	ICC	95 % CI	Cronbach alpha
Focused	2.60 (0.98)	0.76	0.37, 0.88	0.98
The EPA is focused	2.59 (0.99)	0.78	0.39, 0.88	
The EPA is not too broad in scope	2.65 (0.99)	0.77	0.46, 0.90	
The EPA addresses a single activity	2.55 (1.02)	0.70	0.14, 0.84	1.00
Observable	3.62 (0.80)	0.75	0.40, 0.88	
The EPA addresses a visible action	3.60 (0.81)	0.76	0.40, 0.88	
The EPA can be seen	3.63 (0.80)	0.75	0.39, 0.88	0.83
The EPA is observable	3.63 (0.80)	0.75	0.39, 0.88	
Realistic and Generalizable	4.30 (0.57)	0.72	0.08, 0.82	
The EPA is a realistic activity for an internist	4.53 (0.53)	0.61	-0.43, 0.73	0.96
The EPA deals with an activity pertinent to internal medicine.	4.55 (0.53)	0.40	-1.29, 0.56	
The EPA is an activity that an internist would be expected to perform	4.54 (0.59)	0.43	-0.81, 0.65	
The EPA is generalizable to different aspects of medical practice.	4.06 (0.93)	0.78	0.42, 0.89	0.85
The EPA transfers to multiple settings	4.03 (0.95)	0.85	0.63, 0.93	
The EPA is useful across many specialties	4.09 (0.98)	0.78	0.33, 0.87	
Multiple Competencies	4.35 (0.61)	0.77	0.36, 0.88	0.96
The EPA incorporates multiple abilities	4.36 (0.60)	0.76	0.31, 0.87	
The EPA reflects multiple competencies	4.31 (0.70)	0.74	0.36, 0.88	
The EPA requires integration of knowledge, skills, and attitudes	4.39 (0.61)	0.74	0.21, 0.85	0.85
Overall	3.83 (0.44)	0.75	0.001, 0.81	

Another finding was that activities identified as EPAs had higher QUEPA ratings than those identified as OPAs. When asking rotation directors to create EPAs, some created items that reflected OPAs, as they were much more limited in scope or setting. These OPAs may be very commonly observed in some rotation settings, but are unlikely to be generalizable to other settings or contexts. An example of an OPA from our study was “manage hypertension in the peripartum period.” While this is an important skill, it is not likely to generalize to other settings or patient populations. One EPA from our study reads, “Recommend appropriate preventive services or diagnostic interventions based on review of the patient’s current goals of care, prognosis, and evidence based guidelines.” This example illustrates an activity that expands beyond a singular setting or context, to include a potential range of behaviors that would reflect degrees of skill over the course of residency training. Using the QUEPA instrument to rate the quality of EPAs may help programs to avoid choosing items that are overly narrow in scope or content.

The QUEPA ratings reported in this study reflect diverse validity evidence.^{26,27} The main validity evidence for this study is content, which is supported by QUEPA items that were based on existing EPA literature, input from team members with expertise in resident education and scale design, and instrument refinement through the use of iterative pilot testing. Internal structure validity is supported by instrument dimensionality as demonstrated by factor analysis, and excellent inter-rater and internal consistency reliability. Criterion (i.e., relations to other variables) validity is supported by associations between QUEPA ratings and meaningful characteristics of the other variables, including ACGME competency, practice location and activity type.

This study has several limitations. First, it was performed at a single institution, so the findings may not generalize to all settings. Second, we utilized a cross-sectional study design of locally developed EPAs and one set of nationally developed

EPAs; other EPAs will need to be studied in the future. Third, we report only the creation of EPAs and their quality ratings, which represent Kirkpatrick Level 3 (behaviors) and Level 1 (reaction).²⁸ Nonetheless, a systematic review showed that most education research studies report outcomes at the reaction level,²⁹ and behavior level outcomes have been noted to strike the ideal balance between feasibility and rigor.³⁰ Finally, it should be noted that the high inter-rater reliability we reported for the assessment of EPAs was achieved through repeated discussions; similar results may not be obtained elsewhere without a similar degree of attention to the rating process.

This study has important implications for graduate medical education. The Education Redesign Committee of the Alliance for Academic Internal Medicine (AAIM) has recently introduced 16 end-of-training EPAs for internal medicine training.¹⁷ The authors note that these EPAs, while well developed and very appropriate for the end of training, may be overly broad. This statement is supported by our study, which revealed that EOTEPAs had low QUEPA scores in the domains of focus and observable. This finding may reflect that EOTEPAs were developed for summative, end-of-training assessments, and not necessarily for continuous assessment throughout training. Therefore, it is likely that programs will need to develop more narrowly focused EPAs for ongoing assessments of resident performance. The QUEPA tool should help training programs create and identify their own high-quality EPAs. Since the QUEPA contains 15 items with excellent reliability, it would be reasonable, for practical applications, to select a smaller number of items from each of the four QUEPA domains. For the purpose of research, however, it would be advisable to replicate the entire QUEPA, given the potential for factor instability when utilizing an instrument in new educational environments.³¹

This study provides a new instrument for assessing the quality of EPAs written by internal medicine faculty. We are

hopeful that this tool can aid in the development of effective EPAs for evaluating resident progress through their training. Care should be taken when writing outpatient EPAs and those related to medical knowledge, as these EPA categories were of lower quality on average and may be more challenging to create. Our findings have implications for graduate medical education as residency programs shift to EPA-informed evaluation systems. Further research should study the mappings of EPAs to sub-competencies and core competencies, compare EPA related assessment to historical assessment methods, determine learners' perceptions of this new approach, and, most importantly, explore the performance of these EPAs in the actual assessment of resident physicians.

Acknowledgements:

Contributors: None.

Corresponding Author: Jason A. Post, M.D.; College of Medicine, Department of Internal Medicine Mayo Clinic, 200 1st St SW, Rochester, MN 55905, USA (e-mail: post.jason@mayo.edu).

Compliance with Ethical Standards:

Funders: None.

Conflict of Interest: The authors have no conflicts of interest to declare.

REFERENCES

1. **Swing SR.** The ACGME outcome project: retrospective and prospective. *Med Teach.* 2007;29:648–54.
2. **Jones MD Jr, Rosenberg AA, Gilhooly JT, Carraccio CL.** Perspective: Competencies, outcomes, and controversy—linking professional activities to competencies to improve resident education and practice. *Acad Med.* 2011;86:161–5.
3. **Nasca TJ, Phillibert I, Brigham T, Flynn TC.** The next GME accreditation system—rationale and benefits. *N Engl J Med.* 2012;366:1051–6.
4. (ACGME) ACGME. Frequently asked questions: Milestones. 2015.
5. **ten Cate O, Scheele F.** Competency-based postgraduate training: can we bridge the gap between theory and clinical practice? *Acad Med.* 2007;82:542–7.
6. **Ten Cate O.** Nuts and bolts of entrustable professional activities. *J Grad Med Educ.* 2013;5:157–8.
7. **Ten Cate O.** Competency-based education, entrustable professional activities, and the power of language. *J Grad Med Educ.* 2013;5:6–7.
8. **ten Cate O, Snell L, Carraccio C.** Medical competence: the interplay between individual ability and the health care environment. *Med Teach.* 2010;32:669–75.
9. **ten Cate O, Young JG.** The patient handover as an entrustable professional activity: adding meaning in teaching and practice. *BMJ Qual Saf.* 2012;21(Suppl 1):i9–12.
10. **Warm EJ, Mathis BR, Held JD, et al.** Entrustment and mapping of observable practice activities for resident assessment. *J Gen Intern Med.* 2014;29:1177–82.
11. **Jonker G, Hoff RG, Ten Cate OT et al.** A case for competency-based anaesthesiology training with entrustable professional activities: An agenda for development and research. *Eur J Anaesthesiol* 2014.
12. **Aylward M, Nixon J, Gladding S. et al.** An Entrustable Professional Activity (EPA) for handoffs as a model for EPA Assessment Development. *Acad Med.* 2014.
13. **Hauer KE, Kohlwes J, Cornett P, et al.** Identifying entrustable professional activities in internal medicine training. *J Grad Med Educ.* 2013;5:54–9.
14. **Hauer KE, Soni K, Cornett P, et al.** Developing entrustable professional activities as the basis for assessment of competence in an internal medicine residency: a feasibility study. *J Gen Intern Med.* 2013;28:1110–4.
15. **Hauer KE, Ten Cate O, Boscardin C, Irby DM, Iobst W, O'Sullivan PS.** Understanding trust as an essential element of trainee supervision and learning in the workplace. *Adv Health Sci Educ : Theory Pract.* 2013.
16. **Colleges AoAM.** Core entrustable professional activities for entering residency. Association of American Medical Colleges. 2014.
17. **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.* 2014.
18. **Chang A, Bowen JL, Buranosky RA, et al.** Transforming primary care training—patient-centered medical home entrustable professional activities for internal medicine residents. *J Gen Intern Med.* 2013;28:801–9.
19. **Mulder H, Ten Cate O, Daalder R, Berkvens J.** Building a competency-based workplace curriculum around entrustable professional activities: the case of physician assistant training. *Med Teach.* 2010;32:e453–9.
20. **Ten Cate O.** AM last page: what entrustable professional activities add to a competency-based curriculum. *Acad Med.* 2014;89:691.
21. **Landis JR, Koch GG.** The measurement of observer agreement for categorical data. *Biometrics.* 1977;33:159–74.
22. **Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG.** Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform.* 2009;42:377–81.
23. **Epstein RM, Hundert EM.** Defining and assessing professional competence. *Jama.* 2002;287:226–35.
24. **Hawkins RE, Sumption KF, Gaglione MM, Holmboe ES.** The in-training examination in internal medicine: resident perceptions and lack of correlation between resident scores and faculty predictions of resident performance. *Am J Med.* 1999;106:206–10.
25. **Kolars JC, McDonald FS, Subhiyah RG, Edson RS.** Knowledge base evaluation of medicine residents on the gastroenterology service: implications for competency assessments by faculty. *Clin Gastroenterol Hepatol.* 2003;1:64–8.
26. **Cook DA, Beckman TJ.** Current concepts in validity and reliability for psychometric instruments: theory and application. *Am J Med.* 2006;119:166 e7–16.
27. **APA, AERA, and National Council on Measurement in Education.** Standards for educational and psychological testing. Washington, DC: American Education Research Association.; 2014.
28. **Beckman TJ, Cook DA.** Developing scholarly projects in education: a primer for medical teachers. *Med Teach.* 2007;29:210–8.
29. **Reed DA, Cook DA, Beckman TJ, Levine RB, Kern DE, Wright SM.** Association between funding and quality of published medical education research. *Jama.* 2007;298:1002–9.
30. **Shea JA.** Mind the gap: some reasons why medical education research is different from health services research. *Med Educ.* 2001;35:319–20.
31. **Beckman TJ, Cook DA, Mandrekar JN.** Factor instability of clinical teaching assessment scores among general internists and cardiologists. *Med Educ.* 2006;40:1209–16.