UGME Curriculum Renewal 2021

Report | Phase I
Assessment Working Group

***

University of Ottawa, Faculty of Medicine
Undergraduate Medical Education

September 15, 2021

Part 1: Introduction

1. Introduction

In November 2019, the Undergraduate Curriculum Committee (UCC) endorsed the development of a comprehensive curriculum renewal project to include the implementation of the Association of Faculties of Medicine of Canada entrustable professional activities (EPAs) described in the 2016 document entitled "AFMC Entrustable Professional Activities for the Transition from Medical School to Residency" for the class of 2026.

In addition to the implementation of a competency-based medical education program, the Curriculum Renewal Project charter sought to:

- define the characteristics, qualities, professional values, and abilities that define a graduate of the Faculty of Medicine at the University of Ottawa,
- integrate recommendations from strategic working groups (social accountability, inter-professional education, patient partnerships);
- review and revise current approaches to assessment and the use of education technology;
- integrate new curricular content (e.g., anti-racism, ethics education and point of care ultrasound) domains.

The review of assessment approaches is needed. In a survey completed by students and faculty on areas that will be challenges to the success of the Undergraduate Medical Education (UGME) Program over the next 3 years, 43% of students and 70% of faculty respondents identified student assessment strategies. The narrative comments from students focused on assessments as lacking alignment to the curriculum or not promoting understanding or application of knowledge.

Given the strategic directions for curriculum renewal and the view of the students and faculty, the educational leadership of the MD Program supported the formation of a working group to review several aspects of the curriculum including assessment tools and strategies that would support the envisioned changes to the curriculum. A final report that synthesizes the recommendations from all the working groups will be presented to the Curriculum Content Review Committee (CCRC) and the UCC. The CCRC will then identify priorities and create working groups to implement the resulting strategic priorities for curriculum renewal.

2. UGME Assessment Working Group Process

The director, Curriculum and the director, Evaluation (UGME) proposed the formation of an assessment working group to begin to consider tools and strategies for assessing EPAs. A general invitation was sent to a mix of leaders in assessment as well as representatives from

Pre-Clerkship, Clerkship, Society, the Individual and Medicine (SIM), Physician Skills Development/*Développement des aptitudes cliniques* (PSD/DAC), Anatomy, UGME operations staff and students and invited them to participate in this working group. The faculty and administrative staff leaders who elected to participate as members of the working group agreed to the following three-step process.

- Step 1: Working group members reviewed and revised the terms of reference and mandate of the committee.
- Step 2: The working group members identified the need to review the assessment tools currently being used to determine if there are any gaps in how these tools would assess the national EPAs.
- Step 3: Create a series of recommendations with rationale for what changes should be considered to support a revised assessment process.

The inaugural meeting was held in December 2020 at which point a draft terms of reference was presented for consideration. To help support the work of the committee, a review of the literature was conducted using EPA, assessment and UGME plus variations of these terms as search terms. Due to the absence of a key article in this search, a second review was then conducted. These abstracts were then curated by one of the co-chairs of the Assessment Working Group. The titles and abstracts were shared with the working group at the second meeting held in January 2021 and list of the titles appear in Appendix 1.

The second meeting of the working group also finalized the terms of reference and approved the following mandate for the working group:

1. *Complete a gap analysis of the components of the current UGME assessment system in relationship to:*
    - Established MD Program objectives
    - National Entrustable Professional Activities (EPAs).

2. *Define broad recommendations for:*
    a. Creating an integrated assessment system (e.g., longitudinal, programmatic, competency-based, etc.).
    b. Identifying the infrastructure required to enhance the validity of assessment results for the MD Program;
    c. Enhancing the use and frequency of low stakes assessments; and
    d. Identifying data sources for program evaluation of the MD curriculum.

Given the emphasis on mapping the current curriculum to the EPAs, at the third meeting in February 2021, Dr. Michelle Gibson and Ms. Eleni Katsoulas from Queens University discussed

the steps Queens University Medical School underwent to inform their EPA assessment process. This included a mapping of their clerkship assessment tools against the EPAs and CanMEDS roles. In addition, Dr. Craig Campbell presented the list of the first six EPAs that had been approved by the EPA Implementation Working Group. This list also included a description of the knowledge, skills, attitudes, and behaviours (Section 5) required to adequately perform each EPA. The working group members agreed to create subgroups to accomplish the mapping exercise. Each subgroup would focus on a particular assessment tool(s) and would consider it longitudinally across the UG program.

Between March 2021 and May 2021, the Director of Evaluation (UGME), the Director of Curriculum and the Curriculum Renewal Coordinator and the Assessment, Evaluation, and Curriculum (AEC) Office gathered examples of all the assessment tools that were used to assess medical students. For each assessment tool, a subgroup of content experts was recruited to complete the mapping of these assessment tools to the knowledge, skills, attitudes or behavioural descriptions for EPA 1 to 6.

The results of the mapping exercise for EPA 1 to 6 were presented to the Assessment Working Group at its May 2021 committee for consideration.

3. Observations from the mapping exercise

Appendix 2 displays the results of the mapping of the assessment tools used in the MD Program against the first six EPAs. Appendix 3 displays the assessment tools mapped against the EPAs and the Section 5 categories (knowledge, skills, attitudes and behaviours) that define the first six EPAs. Appendix 4 displays the mapping but with each item included in every rating scale.

Based on an analysis of Appendix 2, the first six EPAs are assessed at some point in the UGME curriculum, however, not all assessment tools map to an EPA. The E-portfolio, cased-based learning (CBL), and PSD/DAC rating assessment forms mapped to a small number of EPAs and as seen in Appendix 3, map to an even smaller number of knowledge, skills, attitudes and behaviour descriptions (Section 5). The Objectively Structured Clinical Examination (OSCE), Mini-CEX and General Clerkship rating forms map to a large number of EPAs and Section 5 descriptions. The written exams also map to a large number of EPAs and Section 5 descriptions with the exception of the SIM curriculum. The representative from SIM who helped with the mapping commented that the SIM exams could test more of the Section 5 descriptions but as currently structured the SIM examinations do not do so.

With regards to the Mini-CEX, it was noted that the Mini-CEXs used in the francophone stream included a professionalism rating scale called the P-MEX that is not used in the Anglophone stream.

4. Recommendations

One of the goals of the Assessment Working Group mandate was to provide recommendations around specific themes common in both competency-based assessment and assessing EPAs. In doing so, the themes and recommendations should promote innovation in assessment and not be tied to traditional methods of assessing learners. These themes are identified below and where possible, implications from the mapping exercise will be discussed within the context of the themes. These themes and recommendations are broadly defined and are intended to inform the CCRC. Finally, individual recommendations do not necessarily need to be linked to EPAs or entrustment decisions. Specific details about how these themes and recommendations will be implemented is beyond the scope of this report and will constitute the task of a future assessment implementation group.

Theme 1: Adopt an assessment *for* learning strategy rather than an assessment *of* learning strategy

Assessment is an important driver for student learning hence the phrase "assessment drives learning." Some forms of assessment, however, lend themselves better to learning than other forms of assessment. To make the distinction between how assessment influences learning, the literature describes "assessment of learning" and "assessment for learning." Assessment of learning is typically associated with higher stakes and/or summative assessments. This assessment strategy is prevalent in the current curriculum as evidenced by the number of end-of-unit high stakes exams and student rating forms that are used for pass/fail decisions. Competency-based medical education places an increased reliance on assessment tools that are designed to support learning rather than assess what was learned. Strategies typically used in assessment *for* learning include frequent low stakes exams with feedback, and student assessment forms that include extensive narrative comments rather than just rating scales. These strategies encourage self-learning, self-reflection and provide opportunities for coaching (Schuwirth & Van Der Vleuten, 2011).

Assessment *of* learning can create a tension between students focusing more on studying to achieve high marks on a test, which may lead to unwanted study habits (cramming or rote memorization among others), and the promotion of students focusing on understanding, which aligns with longer-term retention. Even the format of assessment can have an influence on learning and shift the emphasis from achieving a 'pass score' on a test that provides marginal feedback to areas where there are gaps in knowledge to receiving feedback from an assessment as a stimulus for continuous growth and improvement. The following recommendations are designed to address this tension by encouraging assessment *for* learning practices.

Recommendation 1: Review assessment forms in the E-portfolio, CBL, Team-based learning (TBL), and PSD/DAC to ensure they are appropriate for both assessment for learning purposes and the assessment of EPAs for implementation in the 2022-23 academic year.

Rationale: The findings from the mapping exercise revealed that the currently constructed student rating forms for these educational activities did not align well to the assessment of EPAs. The strength of these educational strategies is that they are designed to promote feedback, coaching and self-reflection but the rating instruments that are used do not reflect this purpose. For example, the rating scale in PSD uses pass/fail as the rating scale anchors reflecting a summative rather than formative purpose. Therefore a review of the rating forms for these educational activities should be conducted to ensure the rating scale items align with the principles of assessment *for* learning that narrative comments are encouraged, and in doing so, ensure that these educational activities can contribute data to the achievement of relevant EPAs.

Recommendation 2: Review and enhance the feedback given to students from all high stakes exams.

Rationale: Feedback is considered to be an important driver of learner improvement and is an important aspect of many of the assessment tools within UGME. Unfortunately, the feedback given to students from summative assessments is limited, often consisting of a total score and pass/fail standing. Students are allowed to ask for a review of their exams but even so, the feedback remains limited in terms of promoting learning. The purpose of these assessments is for high stakes summative decisions but the feedback that is given to students should be aligned with an assessment *for* learning approach when feasible. Best practice for providing feedback should be reviewed and adopted for all written exams when deemed feasible.

Recommendation 3: Encourage the adoption of frequent low stakes assessment within courses/units/rotations across all four years of the curriculum.

Rationale: One of the principles of assessment *for* learning is frequent low stakes testing because it encourages the consolidation and retrieval of information (Larsen, Butler, & Roediger, 2008; Michaelsen & Sweet, 2008). It also has the added benefit of encouraging students to study on a regular basis. There are examples of this strategy in some sections of the current curriculum. For example, in Anatomy there are frequent low stakes lab tests and the Foundations Unit has weekly formative quizzes. TBL uses individual and group readiness assurance tests in order to provide immediate feedback around cases.

Teachers within the medical school are increasingly being encouraged to adopt active learning teaching strategies, like flipped classrooms. As part of this teaching strategy, they should also be encouraged to offer more formative quizzes prior to the start of specific sessions.

Theme 2: Longitudinal assessment

Longitudinal assessment is designed to allow monitoring of learners across time. The primary benefit of this type of assessment practice is that it ensures the performance of students is

assessed at specific intervals and thereby allowing the early identification of students who need support and/or improvement.

Recommendation 4: Design, implement and evaluate a progress test strategy that promotes student learning and continuous growth starting in the 2023-24 academic year.

Rationale: There is a need to promote a better balance between high stakes and low stakes assessment in the MD Program. However, there are a number of important challenges to the redesign of current approaches to assessment to facilitate both formative and longitudinal needs. Progress testing is one strategy that has been utilized to promote a longitudinal 'assessment for learning' strategy. Progress testing can be used to chart a student's growth in knowledge and clinical and allows for the provision of detailed feedback (Pugh & Regehr, 2016). Progress tests are based on the complete domain of knowledge required to complete a program and are provided to the entire student body at regular intervals (for example 4 times per year). There have been a number of studies related to the design and implementation of progress testing that summarize their strengths and abilities of these longitudinal assessment in promoting learning (Pugh & Regehr, 2016; Pugh, Touchie, Humphrey-Murto, & Wood, 2015; Wrigley, van der Vleuten, Freeman, & Muijtjens, 2012).

Recommendation 5: Adopt a longitudinal test format to assessments that occur in longitudinal curricula.

Rationale: There are multiple examples of longitudinal curricula in place such as SIM, Interprofessional Education, Clinical Skills Education, Anatomy, Histology, Radiology and Professionalism. This type of curriculum emphasizes topics that are taught over extended periods of time and across multiple years of training. For example, coaches in the e-portfolio program interact with students over the course of an academic year and therefore have opportunities to guide the learner as they acquire skills and knowledge over the year. Despite curricular topics that extend over long periods of time, the assessment tools that are used in all years of the medical school are based on a unit or rotation and focus on short time periods.

Longitudinal assessment moves away from this purely summative end of unit approach to a partially formative approach to testing. It involves administering shorter but more frequent assessments of medical knowledge repeatedly for a period of time, often with explanations of correct answers or feedback. These tests can also be spaced apart in order to test knowledge and skills across time rather than at the end of a specific time period. The philosophy of a longitudinal assessment approach is that through a recurring assessment process, concepts and information are reinforced so that knowledge is better retained and gradually accumulated with gaps in knowledge and skills easily identified early in training.

Given the anticipated inclusion of additional longitudinal curricula (for example, Virtual Care, POCUS, Anti-racism among others) the assessment strategies aligned to these curricula should be equally longitudinal.

Theme 3: Revise existing clinical assessment tools in clerkship to promote the assessment of EPAs

One of the goals of introducing EPAs is to ensure students are ready for the transition to residency thus making clerkship the ideal place to start to introduce EPA-based assessments. Therefore, the recommendations under this theme focused on assessment tools used in clerkship.

Recommendation 6: Review and revise the Mini-CEX form to incorporate assessments of EPAs.

Rationale: The Mini-CEX is a workplace assessment tool that requires observation of a learner as they interact with patients. It is currently used on all of the clerkship rotations and students in all rotations are expected to have two Mini-CEX forms completed per rotation (Anglophone OBGYN requires three). The tool itself consists of a six-item rating scale, a global rating scale and space for comments. As displayed in Appendix 3, there is a high degree of overlap between the Section 5 descriptions of EPAs 1 through 6 when the global rating is compared to the other items on the form. In addition, research conducted on uOttawa Mini-CEXs has shown a high correlation between ratings on the items and the global rating (Humphrey-Murto, Côté, Pugh & Wood, 2017). Given the similarity in ratings and Section 5 descriptions, it is doubtful that the global rating is providing unique information about learners. It is possible therefore to adapt the Mini-CEX currently used in all clinical clerkships to assess one or more EPAs and in doing so, replace the current global rating with an entrustment scale. The logistics of this approach would still need to be determined to decide how best to integrate an EPA (e.g., one EPA / form with six or more EPA assessments) but making this change should reduce some of the redundancy in the Mini-CEX ratings and encourage an increase in the number of observations.

Recommendation 7: Review the educational and administration support of the Mini-CEX

Rationale: When Humphrey-Murto et al. (2017) studied the Mini-CEX that was administered in clerkship one of their findings was that many students were missing data. Furthermore, raters tended to provide similar ratings suggesting that they were not making full use of the rating scale. Given these findings, introducing an entrustment rating or ratings of specific EPAs may require faculty development to raters to encourage the use of the full scale, education to learners to help them interpret low or high entrustment ratings.

In addition, as part of the Mini-CEX, several of the francophone rotations administer a professionalism scale called the P-MEX. How the scale is used by the rotations varies and it is not clear why only the francophone rotations use it. A policy on how the scale is used and standardized is needed.

Recommendation 8: Review the Clerkship general rating forms (Form A) to determine if explicit ratings of EPAs could be included.

Rationale: As shown in Appendix 2 and 3, these rating forms have considerable overlap with other assessment tools used in clerkship. To a large degree that makes sense because these forms are meant to capture a range of existing assessments. However, it may be possible to modify these forms in some manner to provide information that is more consistent with the assessment of the EPAs.

Recommendation 9: Review the OSCE assessments to pilot the inclusion of an entrustment rating for years 2 through 4 and in doing so study how best to incorporate EPAs within an OSCE and study how the information could be used by both learners and the UG program.

Rationale: As shown in Appendix 2 and 3, these rating forms are already covering many of the EPAs and Section 5 descriptions. Given that similarity, the inclusion of an additional entrustment rating scale can be added to the existing assessment tools that are already being used on the OSCE. This approach was used by Queens University on their clerkship OSCEs. An entrustment rating has also been introduced on a trainee OSCE and has shown both good psychometric properties as well as acceptance by raters (Halman, Fu, & Pugh, 2020).

Theme 4: Incorporate a programmatic assessment strategy

The assessment processes supported by the MD Program includes written examinations at the end of each unit; clinical skills assessment through summative OSCE examinations and knowledge assessment combined with end-of-rotation clinical evaluations completed at the end of each core clerkship rotation. There are formative OSCEs that occur throughout the first two years of the MD Program and students have access to practise exams provided to support their learning within the various units and rotations as well to support their preparation for the Medical Council of Canada's LMCC exam. Finally, with the last year (2020/2021) the students initiated an MCQ development strategy that provides assessment questions aligned to the learning objectives for each formal session throughout the first two years of the MD Program.

As seen in the assessment mapping exercise, many of the assessment tools that are currently being used have a high degree of overlap in the EPAs and associated knowledge, skills and attitudes that are being assessed. Given the number of assessment tools being used, and the overlap in what is being assessed, there is an opportunity to consider adopting a programmatic assessment framework to manage the assessment system within the MD Program.

Recommendation 10: Design and implement a programmatic assessment model to comprehensively evaluate the program objectives established for the MD Program.

Rationale: The transition to a competency-based curriculum will require a comprehensive evaluation of competence that no single instrument can measure. The development of a programmatic approach to assessment combines a number of individual assessment activities to come to an accurate judgment of the "competencies being measured" and "compensates for deficiencies through combining several instruments"(Timmerman & Dijkstra, 2017). A programmatic design for assessment within the MD program is important to ensuring there is an evaluation of the competencies that are not easily captured within the EPAs. Formulating a stepwise approach to programmatic assessment will help the MD Program to determine what is and what is not being measured, how multiple individual assessments are aggregated to inform both the learning process of students and the high stakes pass/fail decisions of the Competence Committee, the Promotions Committee and the Student Promotion Executive Committee. Finally, a programmatic design to assessment within the MD Program has the promise of reducing bias in the assessment of complex tasks to ensure there is an appropriate sampling strategy of observations.

5. Implications for implementation

Based on the recommendations of the Assessment Working Group the following section identifies issues relevant to future implementation.

Faculty Development

With the adoption of new assessment tools, increased use of entrustment ratings and the importance of providing valuable feedback including narrative comments, faculty development will be required to ensure raters are using the assessment tools properly. For example, Robinson et al. (2021) found that raters can have difficulty interpreting the anchors used on an entrustment scale. If narrative comments are to be encouraged, raters may need help providing comments that are useful for learning (Dudek, Marks, Wood & Lee, 2008; Halman et al., 2016). It is recommended therefore that a UGME specific faculty development program be designed to include training sessions on new approaches to assessment.

Technology

Elentra is already a key technology piece in UGME and reliance on this platform will increase. The concern from an assessment perspective is that large components of the Elentra Platform are still in development (e.g., exam management) and functionality of the platform to support revisions to current assessments, new assessment strategies or tools, or creating a reporting mechanism that will support oversight of the curriculum is not clear. There have also been concerns raised at the postgraduate level that reporting mechanisms in Elentra are onerous which is a challenge to the expanded use and completion of tools in the Elentra platform. It is also not clear to what degree Elentra supports mobile technology. The Assessment Working Group therefore would like to recommend that the Education Technology Working Group

ensures that technology, whether it is Elentra or another platform, creates efficiencies, supports scholarship, facilitates feedback and should be leveraged to its highest possibilities without creating unnecessary barriers.

Program Evaluation

One of the goals of the Assessment Working Group was to identify data sources that could be used for program evaluation purposes. Obvious data sources include results from exams and ratings but there are other sources of information that are relevant. For example, feedback from students and faculty to determine what works and/or does not work as intended would be of benefit. Historical trends and links to other assessments that are relevant should be considered (i.e., MCC). Stakeholder feedback, including feedback from operations staff, should be included on an ongoing basis to assure the changes that are made reflect what we are trying to achieve.

6. Members of the Assessment Working Group plus subgroups

Dr. Timothy Wood (co-chair)	*PhD, Director, DIME, Director of Evaluation (UGME)*
Dr. Craig Campbell (co-chair)	*MD, FRCPC, FSACME, Medicine, Director Curriculum*
Dr. Michelle Anawati	*MSW, MD, CCFP, Family Medicine, Co-Director Clerkship, Francophone stream*
Dr. Doug Archibald	*PhD, Family Medicine*
Dr. Émilie Braschi	*MD, Family Medicine*
Dr. Isabelle Burnier	*MD, Co-Director Pre-Clerkship, Clinical Skills/ DAC Francophone stream*
Dr. Sophie De Roock	*MD, Ottawa Heart Institute*
France Ferreira	*AEC*
Rohit Gopinath	*Medical student*
Dr. Samantha Halman	*MD, Medicine*
Dr. Susan Humphrey-Murto	*MD, Medicine, DIME*
Sophie Lamontagne	*Medical student*
Éric Larouche	*UG Operations*
Dr. John Leddy	*PhD, Cellular and Molecular Medicine*
Mélanie McCallum	*Medicine Hospital Coordinator*
Dr. Annabelle Pelletier	*MD, Psychiatry Clerkship Director*
Dr. Barb Power	*MD, Medicine, Director, Clinical Skills, Anglophone stream*
Janani Ramamurthy	*Medical student*
Dr. Chris Ramnanan	*PhD, DIME, Director of Practical Exams*
Dr. Gabrielle Weiler	*MD, Pediatrics Clerkship Director*

7. References

Cheung, W. J., Dudek, N., Wood, T. J., Frank, J. R., & Ed, M. A. (2016). Daily Encounter Cards — Evaluating the Quality of Documented Assessments. *J Grad Med Educ*, *8*(4), 601–604.

Dudek, N. L., Marks, M. B., Wood, T. J., & Lee, A.C. (2008). Assessing the quality of supervisors' completed clinical evaluation reports. *Medical Education*, *42*(8), 816–822. https://doi.org/10.1111/j.1365-2923.2008.03105.x

Halman, S., Dudek, N., Wood, T., Pugh, D., Touchie, C., McAleer, S., & Humphrey-Murto, S. (2016). Direct Observation of Clinical Skills Feedback Scale: Development and Validity Evidence. *Teaching And Learning In Medicine*, *1334*(June), 1–10. https://doi.org/10.1080/10401334.2016.1186552

Halman, S., Fu, A. Y. N., & Pugh, D. (2020). Entrustment within an objective structured clinical examination (OSCE) progress test: Bridging the gap towards competency-based medical education. *Medical Teacher*, *42*(11), 1283–1288. https://doi.org/10.1080/0142159X.2020.1803251

Humphrey-Murto, S., Côté, M., Pugh, D., & Wood, T. J. (2017). Assessing the Validity of a Multidisciplinary Mini-Clinical Evaluation Exercise. *Teaching and Learning in Medicine*, *1334*(December), 1–10. https://doi.org/10.1080/10401334.2017.1387553

Larsen, D. P., Butler, A. C., & Roediger, H. L. (2008). Test-enhanced learning in medical education. *Medical Education*, *42*(10), 959–966. https://doi.org/10.1111/j.1365-2923.2008.03124.x

Michaelsen, L. K. & Sweet, M. (2008). The essential elements of team-based learning. *New Directions for Teaching and Learning, 2008*(116), 7–27.

Pugh, D., & Regehr, G. (2016). Taking the sting out of assessment: is there a role for progress testing? *Medical Education*, *50*(7), 721–729. https://doi.org/10.1111/medu.12985

Pugh, D., Touchie, C., Humphrey-Murto, S., & Wood, T. J. (2015). The OSCE progress test – Measuring clinical skill development over residency training. *Medical Teacher*, *00*(00), 1–6. https://doi.org/10.3109/0142159X.2015.1029895

Robinson, T. J. G., Wagner, N., Szulewski, A., Dudek, N., Cheung, W. J., & Hall, A. K. (2021). Exploring the use of rating scales with entrustment anchors in workplace-based assessment. *Medical Education*, 0–3. https://doi.org/10.1111/medu.14573

Schuwirth, L. W. T., & van der Vleuten, C. P. M. (2011). General overview of the theories used in assessment: AMEE Guide No. 57. *Medical Teacher*, *33*(10), 783–797. https://doi.org/10.3109/0142159X.2011.611022

Timmerman, A. A., & Dijkstra, J. (2017). A practical approach to programmatic assessment design. *Advances in Health Sciences Education*, *22*(5), 1169–1182. https://doi.org/10.1007/s10459-017-9756-3

Wrigley, W., van der Vleuten, C. P. M., Freeman, A., & Muijtjens, A. (2012). A systemic framework for the progress test: strengths, constraints and issues: AMEE Guide No. 71. *Medical Teacher*, *34*(9), 683–697. https://doi.org/10.3109/0142159X.2012.704437

Zenisky, A. L., & Hambleton, R. K. (2012). Developing Test Score Reports That Work: The Process and Best Practices for Effective Communication. *Educational Measurement: Issues and Practice*, *31*(2), 21–26. https://doi.org/10.1111/j.1745-3992.2012.00231.x

Appendix 1: Curated list of titles from a literature review that could inform the assessment of EPAs (as of March 2021).

Aulet, T. H., Moore, J. S., Callas, P. W., Nicholas, C., & Hulme, M. (2020). (En)trust me: Validating an assessment rubric for documenting clinical encounters during a surgery clerkship clinical skills exam. *American Journal of Surgery*, *219*(2), 258–262. https://doi.org/10.1016/j.amjsurg.2018.12.055

Bray, M. J., Bradley, E. B., Martindale, J. R., & Gusic, M. E. (2020). Implementing Systematic Faculty Development to Support an EPA-Based Program of Assessment: Strategies, Outcomes and Lessons Learned. *Teaching and Learning in Medicine*, *33(4)*, 434–444. https://doi.org/10.1080/10401334.2020.1857256

Caro Monroig, A. M., Chen, H. C., Carraccio, C., Richards, B. F., Ten Cate, O. & Balmer, D. F. (2021). Medical Students' Perspectives on Entrustment Decision-Making in an EPA Assessment Framework: A Secondary Data Analysis. *Academic Medicine : Journal of the Association of American Medical Colleges*, *96*(8), 1175–1181. https://doi.org/10.1097/ACM.0000000000003858

Carraccio, C., Englander, R., Gilhooly, J., Mink, R., Hofkosh, D., Barone, M. A., & Holmboe, E. S. (2017). Building a Framework of Entrustable Professional Activities, Supported by Competencies and Milestones, to Bridge the Educational Continuum. *Academic Medicine : Journal of the Association of American Medical Colleges*, *92*(3), 324–330. https://doi.org/10.1097/ACM.0000000000001141

Chamberlain, N. R., Sexton, P. S., Hardee, M. R., & Baer, R. W. (2018). Physician-Mentored Patient Rounds to Observe and Assess Entrustable Professional Activities 1 and 2 in Preclinical Medical Students. *The Journal of the American Osteopathic Association*, *118*(3), 199–206. https://doi.org/10.7556/jaoa.2018.039

Colbert-Getz, J. M., Lappe, K., Northrup, M., & Roussel, D. (2019). To What Degree Are the 13 Entrustable Professional Activities Already Incorporated Into Physicians' Performance Schemas for Medical Students? *Teaching and Learning in Medicine*, *31*(4), 361–369. https://doi.org/10.1080/10401334.2019.1573146

Colbert-Getz, J. M., & Shea, J. A. (2021). Three key issues for determining competence in a system of assessment. *Medical Teacher*, *43*(7), 853–855. https://doi.org/10.1080/0142159X.2020.1804540

Curran, V. R., Deacon, D., Schulz, H., Stringer, K., Stone, C. N., Duggan, N., & Coombs-Thorne, H. (2018). Evaluation of the Characteristics of a Workplace Assessment Form to Assess Entrustable Professional Activities (EPAs) in an Undergraduate Surgery Core Clerkship. *Journal of Surgical Education*, *75*(5), 1211–1222. https://doi.org/10.1016/j.jsurg.2018.02.013

D'Agostino, D., & Papa, F. J. (2016). Outcomes-Oriented Medical Training: A Critical Curricular Design Consideration in Developing 21st Century Health Care Professionals. *The Journal of the American Osteopathic Association*, *116*(11), 742–746. https://doi.org/10.7556/jaoa.2016.145

Douglass, K. A., Jacquet, G. A., Hayward, A. S., Dreifuss, B. A., & Tupesis, J. P. (2017). Development of a Global Health Milestones Tool for Learners in Emergency Medicine: A Pilot Project. *AEM Education and Training*, *1*(4), 269–279. https://doi.org/10.1002/aet2.10046

Duggan, N., Curran, V. R., Fairbridge, N. A., Deacon, D., Coombs, H., Stringer, K., & Pennell, S. (2021). Using mobile technology in assessment of entrustable professional activities in undergraduate medical education. *Perspectives on Medical Education*, *10*, 372–377. https://doi.org/10.1007/s40037-020-00618-9

Fazio, S. B., Ledford, C. H., Aronowitz, P. B., Chheda, S. G., Choe, J. H., Call, S. A., Gitlin, S. D., Muntz, M., Nixon, L. J., Pereira, A. G., Ragsdale, J. W., Stewart, E. A., & Hauer, K. E. (2018). Competency-Based

Medical Education in the Internal Medicine Clerkship: A Report From the Alliance for Academic Internal Medicine Undergraduate Medical Education Task Force. *Academic Medicine : Journal of the Association of American Medical Colleges*, *93*(3), 421–427. https://doi.org/10.1097/ACM.0000000000001896

Geraghty, J. R., Ocampo, R. G., Liang, S., Ona Ayala, K. E., Hiltz, K., McKissack, H., Hyderi, A., Ryan, M. S., & Core Entrustable Professional Activities for Entering Residency Pilot Program. (2021). Medical Students' Views on Implementing the Core EPAs: Recommendations from Student Leaders at the Core EPAs Pilot Institutions. *Academic Medicine : Journal of the Association of American Medical Colleges*, *96*(2), 193–198. https://doi.org/10.1097/ACM.0000000000003793

Gruppen, L. D., Ten Cate, O., Lingard, L. A., Teunissen, P. W., & Kogan, J. R. (2018). Enhanced Requirements for Assessment in a Competency-Based, Time-Variable Medical Education System. *Academic Medicine : Journal of the Association of American Medical Colleges*, *93*(3S), S17–S21. https://doi.org/10.1097/ACM.0000000000002066

Hamdy, H. (2018). Medical College of the Future: From Informative to Transformative. *Medical Teacher*, *40*(10), 986–989. https://doi.org/10.1080/0142159X.2018.1498628

Hauer, K. E., Boscardin, C., Fulton, T. B., Lucey, C., Oza, S., & Teherani, A. (2015). Using a Curricular Vision to Define Entrustable Professional Activities for Medical Student Assessment. *Journal of General Internal Medicine*, *30*(9), 1344–1348. https://doi.org/10.1007/s11606-015-3264-z

Hauer, K. E., & Lucey, C. R. (2019). Core Clerkship Grading: The Illusion of Objectivity. *Academic Medicine : Journal of the Association of American Medical Colleges*, *94*(4), 469–472. https://doi.org/10.1097/ACM.0000000000002413

Hawkins, R. E., Welcher, C. M., Holmboe, E. S., Kirk, L. M., Norcini, J. J., Simons, K. B., & Skochelak, S. E. (2015). Implementation of competency-based medical education: Are we addressing the concerns and challenges?. *Medical Education*, *49*(11), 1086–1102. https://doi.org/10.1111/medu.12831

Holmboe, E. S., Sherbino, J., Englander, R., Snell, L., Frank, J. R., & ICBME Collaborators. (2017). A call to action: The controversy of and rationale for competency-based medical education. *Medical Teacher*, *39*(6), 574–581. https://doi.org/10.1080/0142159X.2017.1315067

Holmstrom, A. L., Chia, M. C., O'Brien, C. L., Odell, D. D., Burke, J., & Halverson, A. L. (2020). Entrustable Professional Activity-Based Summative Performance Assessment in the Surgery Clerkship. *Journal of Surgical Education*, *78*(4), 1144–1150. https://doi.org/10.1016/j.jsurg.2020.12.001

Holzhausen, Y., Maaz, A., Marz, M., Sehy, V., & Peters, H. (2019). Exploring the introduction of entrustment rating scales in an existing objective structured clinical examination. *BMC Medical Education*, *19*(1), 319. https://doi.org/10.1186/s12909-019-1736-2

Jonker, G., Booij, E., Otte, W. R., Vlijm, C. M., Cate, O. T., & Hoff, R. G. (2018). An elective entrustable professional activity-based thematic final medical school year: An appreciative inquiry study among students, graduates, and supervisors. *Advances in Medical Education and Practice*, *9*, 837–845. https://doi.org/10.2147/AMEP.S176649

Keeley, M. G., Gusic, M. E., Morgan, H. K., Aagaard, E. M., & Santen, S. A. (2019). Moving Toward Summative Competency Assessment to Individualize the Postclerkship Phase. *Academic Medicine: Journal of the Association of American Medical Colleges*, *94*(12), 1858–1864. https://doi.org/10.1097/ACM.0000000000002830

Krupat, E. (2018). Critical Thoughts About the Core Entrustable Professional Activities in Undergraduate Medical Education. *Academic Medicine: Journal of the Association of American Medical Colleges*, *93*(3), 371–376. https://doi.org/10.1097/ACM.0000000000001865

Lane, J. L., Soep, J. B., & Hanson, J. L. (2018). Narrative Derived From Medical Student Reflection in Action: Lessons Learned and Implications for Assessment. *Academic Pediatrics*, *18*(3), 354–356. https://doi.org/10.1016/j.acap.2017.11.011

Lockyer, J., Carraccio, C., Chan, M.-K., Hart, D., Smee, S., Touchie, C., Holmboe, E. S., Frank, J. R., & ICBME Collaborators. (2017). Core principles of assessment in competency-based medical education. *Medical Teacher*, *39*(6), 609–616. https://doi.org/10.1080/0142159X.2017.1315082

Lomis, K. D., Russell, R. G., Davidson, M. A., Fleming, A. E., Pettepher, C. C., Cutrer, W. B., Fleming, G. M., & Miller, B. M. (2017). Competency milestones for medical students: Design, implementation, and analysis at one medical school. *Medical Teacher*, *39*(5), 494–504. https://doi.org/10.1080/0142159X.2017.1299924

Meyer, E. G., Chen, H. C., Uijtdehaage, S., Durning, S. J., & Maggio, L. A. (2019). Scoping Review of Entrustable Professional Activities in Undergraduate Medical Education. *Academic Medicine: Journal of the Association of American Medical Colleges*, *94*(7), 1040–1049. https://doi.org/10.1097/ACM.0000000000002735

Oudkerk Pool, A., Govaerts, M. J. B., Jaarsma, D. A. D. C., & Driessen, E. W. (2018). From aggregation to interpretation: How assessors judge complex data in a competency-based portfolio. *Advances in Health Sciences Education: Theory and Practice*, *23*(2), 275–287. https://doi.org/10.1007/s10459-017-9793-y .

Pandit, S., Thomas, M. R., Banerjee, A., Angadi, M., Kumar, S., Tandon, A., Shrivastava, T., Bandopadhyay, D., Jamwal, V. D. S., & Basannar, D. R. (2019). A crossover comparative study to assess efficacy of competency-based medical education (CBME) and the traditional structured (TS) method in selected competencies of living anatomy of first year MBBS curriculum: A pilot study. *Medical Journal Armed Forces India*, *75*(3), 259–265. https://doi.org/10.1016/j.mjafi.2018.01.010 .

Pereira, A. G., Woods, M., Olson, A. P. J., van den Hoogenhof, S., Duffy, B. L., & Englander, R. (2018). Criterion-Based Assessment in a Norm-Based World: How Can We Move Past Grades?. *Academic Medicine : Journal of the Association of American Medical Colleges*, *93*(4), 560–564. https://doi.org/10.1097/ACM.0000000000001939

Peters, H., Holzhausen, Y., Boscardin, C., Ten Cate, O., & Chen, H. C. (2017). Twelve tips for the implementation of EPAs for assessment and entrustment decisions. *Medical Teacher*, *39*(8), 802–807. https://doi.org/10.1080/0142159X.2017.1331031 .

Peters, H., Holzhausen, Y., Maaz, A., Driessen, E., & Czeskleba, A. (2019). Introducing an assessment tool based on a full set of end-of-training EPAs to capture the workplace performance of final-year medical students. *BMC Medical Education*, *19*(1), 207. https://doi.org/10.1186/s12909-019-1600-4

Postmes, L., Tammer, F., Posthumus, I., Wijnen-Meijer, M., van der Schaaf, M., & Ten Cate, O. (2021). EPA-based assessment: Clinical teachers' challenges when transitioning to a prospective entrustment-supervision scale. *Medical Teacher*, *43*(4), 404–410. https://doi.org/10.1080/0142159X.2020.1853688

Shewade, H. D., Jeyashree, K., Kalaiselvi, S., Palanivel, C., & Panigrahi, K. C. (2017). Competency-based tool for evaluation of community-based training in undergraduate medical education in India—A Delphi approach. *Advances in Medical Education and Practice*, *8*, 277–286. https://doi.org/10.2147/AMEP.S123840

Schick, K., Eissner, A., Wijnen-Meijer, M., Johannink, J., Huenges, B., Ehrhardt, M., Kadmon, M., Berberat, P. O., & Rotthoff, T. (2019). Implementing a logbook on entrustable professional activities in the final year of undergraduate medical education in Germany—A multicentric pilot study. *GMS Journal for Medical Education*, *36*(6), Doc69. https://doi.org/10.3205/zma001277

Steinemann, S., Gardner, A., Aulet, T., Fitzgibbons, S., Campbell, A., & Acton, R. (2019). American College of surgeons /Association for Surgical Education Medical Student Simulation-based Surgical Skills Curriculum: Alignment with Entrustable Professional Activities. *American Journal of Surgery*, *217*(2), 198–204. https://doi.org/10.1016/j.amjsurg.2018.10.012

Ten Cate, O., Borleffs, J., van Dijk, M., Westerveld, T., & numerous faculty members and students involved in the subsequent Utrecht curricular reforms. (2018). Training medical students for the twenty-first century: Rationale and development of the Utrecht curriculum "CRU+". *Medical Teacher*, *40*(5), 461–466. https://doi.org/10.1080/0142159X.2018.1435855

Ten Cate, O., Graafmans, L., Posthumus, I., Welink, L., & van Dijk, M. (2018). The EPA-based Utrecht undergraduate clinical curriculum: Development and implementation. *Medical Teacher*, *40*(5), 506–513. https://doi.org/10.1080/0142159X.2018.1435856 .

Thompson, L. R., Leung, C. G., Green, B., Lipps, J., Schaffernocker, T., Ledford, C., Davis, J., Way, D. P., & Kman, N. E. (2017). Development of an Assessment for Entrustable Professional Activity (EPA) 10: Emergent Patient Management. *The Western Journal of Emergency Medicine*, *18*(1), 35–42. https://doi.org/10.5811/westjem.2016.10.31479

Wijnen-Meijer, M., Ten Cate, O., van der Schaaf, M., Burgers, C., Borleffs, J., & Harendza, S. (2015). Vertically integrated medical education and the readiness for practice of graduates. *BMC Medical Education*, *15*, 229. https://doi.org/10.1186/s12909-015-0514-z

Appendices 2 to 4 see attached Excel files