

**Briefing Note:**

**Who should be an assessor for CBME workplace assessments (e.g. EPAs)?**

# **Introduction**

The purpose of this document is to provide a recommendation from the Best Practices in Evaluation and Assessment (BPEA) Working Group to PGMEAC regarding the use of faculty and others for competency-based assessments – particularly workplace-based assessments (WBA) such as Entrustable Professional Abilities (EPAs).

Generally, as experienced physicians and teachers, **faculty have a good understanding of the ‘work’ of their specialty, including the needed quality, safety, effectiveness, and efficiency needed**. Part of the faculty appointment process is an review of credentials clinically and suitability for teaching. Fellows, Senior Residents, and co-residents generally have a more limited view of the work and required performance standards.

# **Background**

There has been some research into assessor or rater effects on assessments, though much of the literature refers to studies on expertise in general, not specifically rater or assessor expertise.

There is evidence that experienced and non-experienced raters select and use observational data differently when making judgments and decisions about learner performance (Govaerts, Schuwirth et al. 2011). **Experts paid more attention to context** (e.g. “this looks like a hospital setting, not general practice”) and **generated significantly more interpretations** (e.g. “he is an authoritarian doctor”) **or evaluations** (e.g. “his physical examination skills are very poor”).

Non-experts paid little attention to context and provided description regarding observed performance (e.g. “he asks if this has happened before”). The non-experts demonstrated a “checklist” approach to their assessments. During the think-aloud portion of this study **the experts also demonstrated self-monitoring and reflective remarks** (e.g. “although I’m not sure if I saw this correctly” or “… on the other hand most senior residents do not know how to handle these problems either”). Non-experts did not demonstrate this type of reflective thinking.

The Govaerts study (Govaerts, Schuwirth et al. 2011) also noted that **experts take more time in assessing complex behaviour,** which is consistent with the literature on expertise (Erickson and Lehmann 1996 as quoted in Govaerts, 2011). Novices, on the other hand, focus on scores, regardless of the complexity of the interaction they are observing. **Experts take more time** to observe, gather information, and analyze the information, before arriving at an appropriate score. The **experts were particularly adept at taking context into consideration** (e.g. setting, patient characteristics, etc.) which the novices tended to ignore.

Overall, Govaerts et al suggest that **expert raters appear to have more elaborate and comprehensive mental models of performance**. While the scores between the experts and non-experts were different early on in the observation, after watching the entire interaction (in this case a video) the final scores for the two groups were similar. On the other hand, one rater noted that “If I had not been forced to think aloud, I would have given a 3 (satisfactory), but if I now reconsider what I said before, I want to give a 2 (borderline).” This does suggest that the study intervention itself (think aloud) potentially changed the ratings given by the observers.

One important observation from this study is that the **differences in rater expertise may affect the feedback that raters give to learners.** From the perspective of EPAs feedback is desired so that the learner can improve. As improved quality and specificity of actionable feedback is a goal for CBME, experienced raters are in the best position to provide such feedback.

A study with teachers also noticed that novice supervisors are more observational in their comments (Kerrins and Cushing 2000 as quoted in Govaerts 2011). The non-expert supervisors in this study focused more on discrete aspects of teaching, while the **experienced supervisors focused more on student learning**.

Experts may utilize pattern recognition when observing and judging performance, particularly when they are under time constraints. As a consequence they may provide a more holistic impression of performance, while non-experienced raters may provide more behaviour details (Murphy and Balzer 1986 as quoted in Govaerts 2011)

In another study, Govaerts and colleagues looked at differences between experienced and inexperienced raters (Govaerts, Van de Wiel et al. 2013). They again noted that **experienced raters made significantly more use of task-specific factors in their decisions.** Additionally, the experienced raters were more likely to link these factors to specific learner behaviours. They were also **more likely and to note the impact of these behaviours on both the patient and the outcome of the interaction.** The authors note that these findings are consistent with other research in industrial and organizational psychology, which show that experts are more sensitive to relevant learner behaviours and have a more advanced schema of competent performance (Cardy, Bernardin et al. 1987 as cited in Govaerts 2013).

Research related to rater cognition noted that **as raters become more experienced with the assessment task they have reported developing stable standards**, akin to criterion-related standards, as opposed to normative standards (Berendonk, Stalmeijer et al. 2013 as cited in Gauthier, 2016 #1375).

In a study that examined how experts solve cases it was observed that the expert faculty did not agree on which specific elements were critical, vs. necessary, vs. useful (Gauthier and Lajoie 2014). **Despite solving cases differently, the experts appeared to have a common concept regarding the key elements of clinical reasoning that learners need to demonstrate for a particular case.**

A systematic review of the literature related to rater judgments for the mini-CEX identified additional research related to experienced vs. novice raters (Lee, Brain et al. 2017). This review identified that **experienced raters (faculty) scored students lower than less experienced raters (residents or junior faculty)** (Hill, Kendall et al. 2009, Rogausch, Beyeler et al. 2015 both as cited in Lee 2017). Additionally, even experienced supervisors have more difficulty with assessments when there are multiple dimensions to observe (Tavares, Ginsburg et al. 2016).

# **Discussion**

Competency-Based Medical Education (CBME) is a relatively new concept in medical education. Many experienced faculty are not yet fully conversant with components of CBME. For example, they may lack understanding of key concepts such as entrustment (Kogan, Conforti et al. 2015) or coaching for performance.

Becoming an expert assessor likely **requires deliberate practice in observing and scoring learners** as well as being provided with feedback on one’s performance as an assessor (Ericsson 2004, Ericsson 2008, Ericsson and Pool 2016)

Given the evidence, it seems that using faculty as assessors for CBME workplace assessments such as EPAs is appropriate whenever possible. All assessors need orientation, faculty development in the actual assessment activities, and feedback on their assessment skills.

A balanced approach that develops, supports, and utilizes both experts (faculty) and novices (senior residents, fellows, co residents) is reasonable. Providing oversight and feedback to the novice assessors would assist them in improving their abilities as assessors.

Appendix – DRAFT Guidelines for the selection of Assessors/Raters in Assessment of Residents



PGME Guidelines for the Selection of Assessors for CBME workplace assessments (e.g. EPAs) of Residents

1. All assessors need orientation, faculty development in the assigned assessment activities, and monitoring and feedback on their assessment skills.
2. The Residency Program Committee (RPC) needs to develop a balanced approach in their Assessment Plan to the selection who will be assessors for CBME workplace assessments (e.g. EPAs) including:
   1. Physician faculty with a faculty appointment will be the assessors for CBME workplace assessments, such as EPAs, whenever appropriate and possible.
   2. Others can act as assessors for CBME workplace assessments, such as EPAs, when deemed appropriate to the assessment activity.
   3. Other assessors for CBME workplace assessments, such as EPAs, include a wide range of clinical team members, clinical associates, clinical fellows, senior residents and co-residents.
   4. The majority (i.e. > 50%) of each of the CBME workplace assessments, such as EPAs, must be completed by appointed faculty, except where the RPC determines otherwise in their Assessment Plan (e.g., Multi Source Feedback assessments).
   5. There needs to be a monitoring of the proportion of assessments, for each resident, completed by appointed faculty.
   6. There needs to be a monitoring of assessors at different locations, sites, rotation etc. to ensure residents are benefiting from appointed faculty expertise as assessors.
3. PGME will support RPCs in their monitoring, oversight and feedback to assessors including:
   1. Establishing automated systems to monitor the proportion of appointed faculty completing assessments.
   2. Providing oversight and feedback to assessors.
   3. Assisting assessors to improve their abilities.
4. Departments, divisions and programs are encouraged to support faculty as assessors through assessor orientation, faculty development, awards, and other available formal and informal recognitions systems.

# References

Berendonk, C., R. Stalmeijer and L. W. T. Schuwirth (2013). "Expertise in performance assessment: assessors' perspectives." Advances in Health Science Education 18: 559-571.

Cardy, R. L., H. J. Bernardin, J. G. Abbott, M. P. Senderak and K. Taylor (1987). "The effects of individual performance schemata and dimension familiarization on rating accuracy." Journal of Ocupational Psychology 60: 197-205.

Erickson, S. S. and A. C. Lehmann (1996). "Expert and exceptional performance: Evidence of maximal adaptatioin to task constraints." Annual Review of Psychology 47: 273-305.

Ericsson, A. (2008). "Deliberate practice and acquision of expert performance: a general overview." Academic Emergency Medicine 15(11): 988=994.

Ericsson, A. and R. Pool (2016). Peak: Secrets from the new science of expertise. NY, NY, Houghton Mifflin Harcourt Publishing Co.

Ericsson, K. A. (2004). "Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains." Academic Medicine 79(10): S70-81.

Gauthier, G. and S. P. Lajoie (2014). "Do expert clinical teachers have a shared understanding of what constitutes a competent reasoning performance in case-based teaching?" Instructional Science 42(4): 579-594.

Govaerts, M. J. B., L. W. T. Schuwirth, C. P. van der Vleuten and A. M. M. Muijtjens (2011). "Workplace-based assessment: effects of rater expertise." Advances in Health Science Education 16: 151-165.

Govaerts, M. J. B., M. W. J. Van de Wiel, L. W. T. Schuwirth, C. P. M. Van der Vlenten and A. M. M. Muijtjens (2013). "Workplace-based assessment: raters’ performance theories and constructs." Advances in Health Science Education 18: 375-396.

Hill, F., K. Kendall, K. Galbraith and J. Crossley (2009). "Implementing the undergraduate mini-CEX: A tailored approach at Southampton University." Medical Education 43: 326-334.

Kerrins, J. A. and K. S. Cushing (2000). "Taking a second look: Expert and novice differences when observing the same classroom teaching segment a second time." Journal of Peronnel Evaluation in Education 14(1): 5-24.

Kogan, J. R., L. N. Conforti, E. Bernabeo, W. Iobst and E. Holmboe (2015). "How faculty members experience workplace-based assessment rater training: a qualitative study." Medical Education 49(7): 692-708.

Lee, V., K. Brain and J. Martin (2017). "Factors Influencing Mini-CEX Rater Judgments and Their Practical Implications: A Systematic Literature Review." Academic Medicine 92(6): 880-887.

Murphy, K. R. and W. K. Balzer (1986). "Systematic distortions in memory-based behaviour ratings and performance evaluation. Consequences for rating accuracy." Journal of Applied Psychology 71: 39-44.

Rogausch, A., C. Beyeler, S. Montagne, P. Jucker-Kupper, C. Berendonk, S. Huwendiek, A. Gemperli and W. Himmel (2015). "The influence of students' prior clinical skills and context characteristics on mini-CEX scorees in clerkships - A multilevel analysis." BMC Medical Education 15: 208.

Tavares, W., S. Ginsburg and K. Eva (2016). "Selecting and Simplifying: Rater Performance and Behavior When Considering Multiple Competencies." Teaching & Learning in Medicine 28(1): 41-51.