

## 35 How Do We Effectively Measure the Milestones?

*Hauff SR, Hopson LR, Stansfield B, Perry MA, Santen SA/University of Michigan, Ann Arbor, MI*

**Background:** As emergency medicine (EM) moves to evaluation and reporting of milestones, care must be taken to understand validity issues with assessment tools. The EM milestone evaluation form is a global evaluation with progressive milestones representing levels of expertise. Each level is behaviorally anchored to encourage a more precise behavioral assessment, however the concern is that faculty are continuing to make gestalt global assessments rather than employing the behavior anchors.

**Objectives:** The purpose of this study was to compare assessments completed by faculty using the standard form with milestones in order and a form where the milestones were randomized, forcing faculty to use the behavior anchors for assessment.

**Methods:** Nine of the milestones were evaluated. For the randomized form, a random number generator was used to create the order in which the level 1-4 skills were listed. Each faculty completed both forms for each resident. Correlation between the forms was calculated using Intraclass Correlation Coefficients.

**Results:** 20 residents were evaluated by 34 faculty. Inter-rater reliability for the standard evaluation was near-perfect (0.96-0.99) indicating that faculty were assigning a global gestalt rating and not making expert assessment judgments. Inter-rater reliability for the randomized form was poor (-0.05-0.68). Rating variance between the ordered and randomized was the same for each milestone evaluated.

**Conclusions:** Given these results, it is likely that by tethering these discrete skills to a developmental continuum, performance information is actually lost. This suggests that our current evaluation tool is not being utilized correctly, and is probably not measuring the discrete behaviors we are interested in. Treating the milestones as a continuum may actually be taking away information, as faculty are not thoughtfully rating residents and instead place them at the point where they “should” be based on post-graduate year.