

Programmatic Assessment: Medical Teacher. 2012;34:205-14

- **Any single assessment data point is flawed**
 - There is no magic bullet
 - Single assessments do not capture growth
 - Individual performance is *highly* context dependent/influenced such that multiple sampling (observations) over time are needed
- **Standardized assessment can have validity built-in (Sources of Validity evidence: Content, Response process, Int structure, Reln to other variables, Consequences)**
 - E.g.: In-training exams
 - Content defined by discipline
 - Rigorous item-development processes
 - Controlled scoring/administration procedures
 - Prior studies link to certification exam success
- **Validity of non-standardized assessment rests with users, less so in instruments**
 - Much of real-world Workplace Assessment
 - Even Milestones are not fully standardized
 - It's not the form, it's the user
 - Users need to be trained
 - Users need to understand their role
 - Can't just hand them a form to complete
 - Without understanding their roles, assessment trivialized
 - "Straight-line 5"; no comments
- **Know the stakes (low vs. high)**
 - All assessments have stakes
 - Low Stakes
 - Limited consequences for promotion, certif, etc.
 - Can be a single data point
 - *For Learning*
 - High Stakes
 - Significant consequences
 - MUST be based on multiple sources of info (including single-data points—*entirety of the record*)
 - Exception: Mastery task (CPR)
 - Will entail judgment: role of teacher/judge may be in conflict
- **Assessment drives learning**
 - Can lead to undesirable behaviors
 - Studying old exams; prep guides; sharing OSCE cases
 - Should generate meaningful fb to learner
 - Both quantitative and qualitative (descriptive)
 - Foster desirable learning activities
 - Low stakes should be as rich as possible
 - Goal is to foster engagement/learning
- **Expert judgment imperative**
 - Clinical competency committees
 - Interpreting assessments requires judgment
 - Synthesis requires time, reflection, context
 - Entirety of the record
 - Overcoming bias
 - Random (hawks/doves): sampling strategies
 - Systematic: procedural measures around decisions
 - Qualitative is NOT subjective
 - Can be rich source to guide decisions