Pediatrics Milestones and Meaningful Assessment

Translating the Pediatrics Milestones into Assessment Items for use in the Clinical Setting

Ann Burke
Susan Guralnick
Patty Hicks

Jeanine Ronan
Dan Schumacher
Alan Schwartz
Rebecca Tenney-Soeiro
Conflict of Interest

There are no industry conflicts of interest to declare for any workshop presenters.
Workshop Outcomes

- Increase participant familiarity with construction of behavioral or observable items for assessment
- Impart concepts of inter-rater reliability, construct validity, rater training and calibration and other key assessment principles as they relate to creating milestone assessment items
- Develop appreciation for the context specificity involved in observing and scoring milestone assessment instruments
Workshop Timeline

Pediatrics Milestones, assessment, learner perspective, teacher/observer Perspective, item writing

Item writing for Pediatrics Milestones

Review of group item writing and Pediatrics Milestones background

Video scoring using newly developed items

Groups report out on their scoring

Discussion
The Pediatrics Milestones

History

- Work commissioned by the ABP and the ACGME
- Led by Carol Carraccio and written by the Pediatrics Milestones Working Group
- Input by APPD membership, with Personal and Professional Development “competency” created
The Pediatrics Milestones

Aim

- To provide descriptions of behaviors that describe learner outcomes
- Allow observers to identify level of learner along a developmental spectrum
- Allow learners to receive feedback regarding their position along the path to becoming a competent physician
The Pediatrics Milestones

*Design*

- Conceptual frameworks created through search of literature for evidence of developmental progression
- Theoretical framework and background provided – establishes grounding for proposed achievement levels for “milestones” for each sub-competency
The Pediatrics Milestones

Current Version

- First iteration published on [ABP website](#)

- Faculty development resource

- Resource for developing items with behavioral anchors, global items

- Resource for developing assessment instruments with clusters of items pertinent to meaningful clinical performance assessment
If Resident Assessment could be so easy...
Validity Concepts

- All validity is Construct Validity which requires multiple sources of evidence. Construct validity has multiple facets.

- Validity refers to the evidence presented to support or refute the meaning or interpretation assigned to assessment results.

- Assessments are not valid or invalid; rather the scores or outcomes of assessment have more or less evidence to support a specific interpretation.

Validity Evidence Sources

- Content
- Response
- Internal structure
- Relationship to other variables
- Consequences
Content Evidence

CONTENT OVER or UNDER REPRESENTATION

Are the standardized patients all portraying the content domains they are supposed to be portraying?

Do the areas “tested” map to domains that are meant to be tested?
So now that we know the Sources of Validity...

- What are some problems or threats to validity?
- We can think of some examples from:
  - WRITTEN TESTS
  - PERFORMANCE EXAMINATIONS (OSCEs)
  - RATINGS OF CLINICAL PERFORMANCE
So now that we know the Sources of Validity...

- What are some problems or threats to validity?
- We can think of some examples from:
  - WRITTEN TESTS
  - PERFORMANCE EXAMINATIONS (OSCEs)
  - RATINGS OF CLINICAL PERFORMANCE
Ratings of Clinical Performance Example

- A resident who has had two people directly observe him performing a physical exam during his residency.

- The faculty all value and weigh various characteristics in their own way; one faculty member rates all males higher than females, the other one gives everyone 5/5 because he “likes them all.”

- Halo effect (rater ignores the traits to be rated and treats all traits as one).
Threats to Validity

- Any factors that interfere with proposed interpretation of assessment scores or ratings threaten validity

- Two types in particular
  - Construct Under-Representation
  - Construct-irrelevant Variance
Construct Under-Representation

- Refers to under sampling in a content domain
- Using too few items, cases or clinical performance observations to adequately generalize to that domain.


- Ratings of Clinical Performance
  - Too few observations of clinical behavior
  - Too few independent raters
Construct-Irrelevant Variance

- Variables that systematically (rather than randomly) interfere with the ability to meaningfully interpret scores or ratings.


- Ratings of Clinical Performance
  - Inappropriate rating items
  - Rater bias
  - Systematic rater error: Halo, severity, leniency
Reliability

- Reliability is a major source of validity evidence for assessments.
- Reliability refers to the reproducibility of assessment data or scores over time or occasions.
- The largest threat to reproducibility of clinical ratings is rater inconsistency or low interrater reproducibility.

End-User Buy-In and Understanding

- Cognitive Interviews with residents
- Cognitive interviews with faculty
Key Guiding Principle and Conceptual Consideration

Make the Pediatrics Milestones learner-centered

Study Objective:

Determine how pediatric residents understand, interpret, and respond to the Pediatrics Milestones.
Methods

- 48 cognitive interviews with peds residents
- Cincinnati Children’s and Dayton Children’s
- 6 interviews with each of 8 Pediatrics Milestones
- Chosen for range of competency domain, complexity, length, and primary author
- Transcription and inductive coding
Residents demonstrate sufficient, often deep, understanding of the Pediatrics Milestones

Residents use their own experiences to understand the Pediatrics Milestones

Residents are comfortable being placed before the end of the developmental continuum of the Pediatrics Milestones

The Pediatrics Milestones serve as roadmaps for learning
Challenges in Real Time Observations

- TIME
- TIME
- TIME

Never enough time
Faculty: The multi-tasker

- Clinical Observer
- Listener
- Observer/Director (Redirector) of Rounds
- Teacher
Role as Redirector/Teacher

- Question the learner and allow them to show knowledge or leave them alone and don’t expose potential gaps?
- Rounds are really about generating great questions, not necessarily what is already known.
- If you prompted a learner where does that fall in the scoring rubric?
Faculty Development

- Again-TIME
- Need for those who can lead the development
- How do we direct faculty around the idea of changing rounds to raise great questions
- How do we all agree on interpretation of observational/feedback tools
Item writing principles

- Behaviors, not evaluations
  - Describing what the resident does
  - Distinct from determining how good, bad, advanced, etc. those behaviors are

- Observed, not intuited
  - Relying on verifiable observations increases reliability
  - Thoughts are hard to observe until they manifest as actions
Item writing principles

- Simple, not complex
  - If the milestone involves multiple dimensions of behavior, you may need multiple items.

- Contextually-appropriate and consistent
  - If the behavior is only sensible in particular contexts, situate the assessment in those contexts

- Under the learner’s control
  - Item should reflect the learner’s behavior, not downstream outcomes that are influenced by factors out of learner’s control