Milestone Ratings and Supervisory Role Categorizations Swim Together, but is the Water Muddy?

Daniel J. Schumacher, MD, MEd¹; Kathleen W. Bartlett, MD²; Sean P. Elliott, MD³;
Catherine Michelson, MD, MMSc⁴; Tanvi Sharma, MD, MPH⁵; Lynn C. Garfunkel,
MD⁶; Beth King, MPP⁷; Alan Schwartz, PhD⁸; and the APPD LEARN CCC Study Group

Affiliations

¹Daniel J. Schumacher, MD, MEd; associate professor, Department of Pediatrics;
Cincinnati Children's Hospital Medical Center/University of Cincinnati, Cincinnati, Ohio

²Kathleen W. Bartlett, MD; associate professor and associate program director,
Department of Pediatrics; Duke University, Durham, North Carolina

³Sean P. Elliott, MD; professor, associate chair, and program director, Department of Pediatrics; University of Arizona, Tucson, Arizona

⁴Catherine Michelson, MD, MMSc; assistant professor and pediatric program director, Department of Pediatrics, Boston University School of Medicine, Boston, Massachusetts

⁵Tanvi Sharma, MD, MPH; assistant professor and associate program director,
Department of Medicine; Boston Children's Hospital and Harvard Medical School,
Boston, Massachusetts

1

2 ⁶Lynn C. Garfunkel, MD; associate program director, Department of Pediatrics;

3 University of Rochester, Rochester, New York

4

5 ⁷Beth King, MPP; research project manager, Association of Pediatric Program Directors

6 (APPD) Longitudinal Educational Assessment Research Network (LEARN), McLean,

7 Virginia

8

9 ⁸Alan Schwartz, PhD; Michael Reese Endowed Professor of Medical Education and

10 Associate Head, University of Illinois at Chicago Department of Medical Education,

11 Research Professor, University of Illinois at Chicago Department of Pediatrics, Chicago,

12 Illinois, and Director, Association of Pediatric Program Directors Longitudinal

13 Educational Assessment Research Network, McLean, Virginia

14

15 **Members of the APPD LEARN CCC Study Group not named above:**

16 Michelle Barnes, MD; associate professor, associate head of education, and associate

17 pediatric program director, Department of Pediatrics, University of Illinois at Chicago,

18 Chicago, Illinois

19

20 Natalie Burman, DO, MEd, assistant professor and associate program director,

21 Department of Pediatrics; Naval Medical Center San Diego, San Diego, California

22

1 Sharon Calaman, MD; associate professor and pediatric program director, Department of
2 Pediatrics, Drexel University College of Medicine, St. Christopher's Hospital for
3 Children, Philadelphia, Pennsylvania

4

5 John G. Frohna, MD, MPH; professor, Departments of Pediatrics and Internal Medicine;
6 University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin

7

8 Caren Gellin, MD; assistant professor and associate program director, Department of
9 Pediatrics; University of Rochester, Rochester, New York

10

11 Kathleen Gibbs, MD; was an assistant professor and associate program director,
12 Department of Pediatrics; Icahn School of Medicine at Mount Sinai, New York at the
13 time of this study. She is now assistant professor, Department of Pediatrics, The
14 Children's Hospital of Philadelphia and The University of Pennsylvania, Philadelphia,
15 Pennsylvania

16

17 Javier Gonzalez del Rey, MD, MEd; professor, Department of Pediatrics; Cincinnati
18 Children's/University of Cincinnati, Cincinnati, Ohio

19

20 Su-Ting T. Li, MD, MPH; professor, vice chair of education and pediatric program
21 director, Department of Pediatrics, University of California Davis, Sacramento,
22 California

23

1 Jon F. McGreevy, MD, MSPH; assistant professor and associate program director,
2 Department of Pediatrics; Phoenix Children's Hospital/University of Arizona, Phoenix,
3 Arizona

4

5 Sue Poynter, MD, MEd; associate professor and pediatric program director, Department
6 of Pediatrics, Cincinnati Children's Hospital Medical Center/University of Cincinnati,
7 Cincinnati, Ohio

8

9 Shannon E. Scott-Vernaglia, MD; assistant professor, Department of Pediatrics;
10 Massachusetts General Hospital and Harvard Medical School; Boston, Massachusetts

11

12 Daniel Sklansky, MD; assistant professor and associate program director, Department of
13 Pediatrics, University of Wisconsin School of Medicine and Public Health, Madison,
14 Wisconsin

15

16 Lynn Thoreson, DO; assistant professor and associate program director, Department of
17 Pediatrics, The University of Texas at Austin Dell Medical School, Austin, Texas

18

19

20 The views expressed in this article are those of the author(s) and do not necessarily
21 reflect the official policy or position of the Department of the Navy, Department of
22 Defense, or the U.S. Government.

23

1 **Corresponding Author**

2 Daniel J. Schumacher, MD, MEd

3 Assistant Professor, Emergency Medicine

4 Cincinnati Children's

5 3333 Burnet Avenue

6 MLC 2008

7 Cincinnati, Ohio

8 USA

9 O: 513-803-2639

10 e-mail: daniel.schumacher@cchmc.org

11

12 **Key Words:** clinical competency committee, milestone-based assessment, graduate

13 medical education, entrustment

14 **Running Title:** Relation between milestone ratings and supervision decisions

15 **Abstract Word Count:** 246

16 **Main Text Word Count:** 3046

17 **Acknowledgements:** This study was provided in-kind support from the Association of

18 Pediatric Program Directors Longitudinal Educational Assessment Research Network

19 **Conflicts of Interest:** None

1 **Abstract**

2 Purpose: This single specialty, multi-institutional study aimed to determine: 1) the
3 association between milestone ratings for individual competencies and average milestone
4 ratings (AMRs) and 2) the association between AMRs and recommended supervisory
5 role categorizations made by individual clinical competency committee (CCC) members.

6 Methods: During the 2015-16 academic year, CCC members at 14 pediatric residencies
7 reported milestone ratings for 21 competencies and recommended supervisory role
8 categories (may not supervise, may supervise in some settings, may supervise in all
9 settings) for residents they reviewed. An exploratory factor analysis of competencies was
10 conducted. The associations between individual competencies, the AMR, and
11 supervisory role categorizations were determined by computing bivariate correlations.
12 The relationship between AMRs and recommended supervisory role categorizations was
13 examined using an ordinal mixed logistic regression model.

14 Results: 68/155 CCC members completed both milestone assignments and supervision
15 categorizations for 451 residents. Factor analysis of individual competencies controlling
16 for clustering of residents in raters and sites resulted in a single-factor solution
17 (cumulative variance 0.75). All individual competencies had large positive correlations
18 with the AMR (correlation coefficient: 0.84-0.93), except for two professionalism
19 competencies (Prof1: 0.63 and Prof4: 0.65). When combined across training year and
20 time points, the AMR and supervisory role categorization had a moderately positive
21 correlation (0.56).

1 Discussion: This exploratory study identified a modest correlation between average
2 milestone ratings and supervisory role categorization. Convergence of competencies on a
3 single factor deserves further exploration, with possible rater effects warranting attention.

4

5

1 **What's New**

2 This exploratory study suggests it may be possible to generate a picture of resident
3 performance by measuring only one or two aspects of that performance. However,
4 additional considerations exist, including individual competency uniqueness and the role
5 of rater effects.

6

7

1 Milestones and entrustment have become a major focus of work-based assessment efforts
2 internationally in recent years.¹⁻¹¹ All residency programs accredited by the
3 Accreditation Council for Graduate Medical Education (ACGME) are required to report
4 milestones for their residents biannually.¹ Additionally, many residency programs see
5 the value in assessing residents using an entrustment framework, the most common of
6 which places focus on the extent to which residents can be entrusted to perform
7 foundational tasks of the profession and the amount of supervision they need to safely do
8 so.^{3,4}

9
10 With both assessment frameworks (milestones and entrustment) co-existing in
11 contemporary popularity in work-based assessment, the relationship between these
12 frameworks is important to elucidate. This is especially important because some
13 programs use one as an indirect measure for the other without making assessment
14 decisions about milestones and entrustment separately.¹² Efforts to explore the
15 association between milestone decisions and entrustment decisions is almost non-
16 existent. Li and colleagues have considered the relationship between milestones and
17 graduation from residency (a type of entrustment decision), but focused solely on the
18 completion of training and did not consider whether residents were truly ready to
19 graduate.¹³ To further explore the association of milestones- and entrustment-based
20 resident assessment, this single specialty, multi-institutional study sought to determine: 1)
21 the association between individual competencies and average milestone ratings (AMRs),
22 and 2) the association between AMRs and recommended supervisory role categorizations
23 made by individual clinical competency committee (CCC) members.

1

2 **Methods**

3 Study Setting

4 This multi-site longitudinal prospective observational cohort study was conducted during
5 the 2015-2016 academic year. Fourteen pediatric residency programs (Table 1) in the
6 Association of Pediatric Program Directors (APPD) Longitudinal Educational
7 Assessment Research Network (LEARN) participated.

8

9 Data Collection

10 All CCC members and categorical pediatrics residents at study sites were considered
11 eligible. For feasibility purposes, site leads were asked to prospectively recruit a
12 convenience sample of CCC members and pediatric residents given the large numbers of
13 both at some sites. If not including all residents, programs were asked to select residents
14 from the full range of performance based on previous performance. We did not ask
15 programs what sampling strategy they employed but the data suggest nine programs
16 sampled all residents, three programs chose to sample top/bottom residents, and the
17 strategy of the remaining two programs cannot not be determined.

18

19 CCC members serve on a committee that reviews and makes summative assessment
20 decisions based on frontline performance level assessment data that are gathered over a
21 specified period of time. These groups are required to report milestone levels to the
22 ACGME at the mid-point and end of the academic year, so the data reviewed are
23 typically from the first and then second halves of the academic year. We sought to

1 collect data for this study at those points in time when CCC biannual reviews were
2 already occurring. CCC members may or may not have experience with residents
3 clinically. Sometimes they are reviewing assessment data for residents they do not know
4 at all and sometimes they are reviewing assessment data for residents they have worked
5 with on several occasions.

6
7 During the Fall and Spring milestone reporting periods of the 2015-2016 academic year,
8 individual CCC member study participants reported two sets of summative assessment
9 decisions for the residents they personally reviewed. First, they reported the ratings they
10 made for each of the 21 ACGME reporting competencies for each resident. In pediatrics,
11 4-5 milestones levels exist for each competency. These 21 competencies all fall within
12 the six ACGME competency domains of patient care (e.g. information gathering, clinical
13 reasoning), medical knowledge (e.g., evidence-based medicine), professionalism (e.g.
14 professional identity formation, trustworthiness), interpersonal and communication skills
15 (e.g., communication across a range of socioeconomic and cultural backgrounds,
16 emotional intelligence), practice-based learning and improvement (e.g. quality
17 improvement, response to feedback), and systems-based practice (e.g., care coordination,
18 inter-professional teamwork). Consistent with allowances when reported to the ACGME,
19 participants were allowed to indicate that a resident fell half way between two levels.

20 Second, they provided a recommended supervisory role categorization for each resident.
21 Six supervisory role category choices were presented: 1) may serve in a supervisory role
22 as a resident in ALL settings; 2) may serve in a supervisory role as a resident in ALL
23 settings, but is just above the borderline/marginal mark for serving in this role; 3) may

1 serve in a supervisory role as a resident in SOME settings; 4) may serve in a supervisory
2 role in SOME settings, but is just above the borderline/marginal mark for serving in this
3 role; 5) may not serve in a supervisory role as a resident; or 6) unable to determine.

4 These categories were developed, reviewed, and edited by a group of 12 residency and
5 medical education research leaders through an iterative process prior to administration.

6
7 Our focus on supervisory role categorization arises from the pediatric graduate medical
8 education community in the United States embracing a view of entrustment that includes
9 important entrustment decisions that happen during training¹⁴⁻¹⁶ in addition to those
10 important for making decisions about readiness to practice outside training.^{3-5,17} These
11 include readiness to serve as an intern, readiness to serve without an onsite supervisor
12 immediately available, and readiness to supervise others.¹⁴⁻¹⁶ These entrustment
13 inferences have been defined by the Pediatric Milestone Assessment Collaborative, an
14 effort of APPD LEARN, the American Board of Pediatrics, and the National Board of
15 Medical Examiners. As noted, this study focuses on the final one of these: readiness to
16 supervise others.

17
18 One program did not have a review process where individual CCC members were
19 assigned residents. For this program, the program director reported the consensus
20 decisions of the group.

21
22 Current CCC members and all categorical pediatrics residents were considered eligible
23 participants at each site. As noted previously, site leads were asked to prospectively

1 recruit a convenience sample of CCC members and pediatric residents given the large
2 numbers of both at some sites. If not including all residents, programs were asked to
3 select residents from a representative range of performance based on previous
4 performance.

6 Data Analysis

7 To try to identify groupings of competencies that might have stronger or weaker
8 associations with supervisory ratings, we first conducted an exploratory factor analysis of
9 competencies to identify whether milestone ratings may reflect a smaller number of
10 underlying latent factors (for example, the 6 ACGME competency domains within which
11 the pediatrics milestones were defined). We were next interested in how these factors co-
12 varied with supervisory role categorizations to determine which variables may be most
13 predictive for this categorization. Because each learner's milestones are clustered in
14 raters in sites, we first regressed milestone ratings on random effects of rater and site, and
15 then performed the factor analysis on the residuals (the ratings excluding the contribution
16 of rater and site). As a sensitivity analysis, we also performed factor analysis on the
17 (partial) correlation matrix of milestone ratings with a fixed effect of rater partialled out,
18 and a factor analysis on the raw ratings without controlling for clustering.

19
20 Extraction was performed using maximum likelihood methods¹⁸, with the number of
21 factors to extract based on parallel analysis.¹⁹ Interpretation was facilitated by an oblique
22 rotation (oblimin), allowing for correlation among factors. Factor analysis was
23 conducted using R 3.3²⁰, and the nFactors²¹ and GPArotation²² packages. We included

1 both Fall and Spring time points in the factor analysis, treating each assignment of
2 milestones as independent. In addition to the factor analysis of the complete data set, we
3 conducted three additional exploratory factor analyses of the PGY1, PGY2, and PGY3
4 data separately, and compared the factor structures.

5
6 An Average Milestone Rating (AMR) was calculated for each resident by averaging all
7 21 milestone level assignments. Although three competencies (SBP1, SBP3, PC4) have
8 only four defined milestone levels, the survey instrument presented options up to 5 for all
9 competencies, and many sites reported level 4.5 and 5 milestone assignments for these
10 competencies (that is, they used a full 5 point scale). Accordingly, all milestones were
11 simply averaged; in a sensitivity analysis, we excluded those three competencies, which
12 did not affect the pattern of findings.

13
14 We examined the associations between individual competencies, the AMR, and
15 supervision decisions by computing bivariate correlations among each pair of variables.

16
17 We also examined the relationship between AMRs and recommended supervisory role
18 categorizations (excluding 8 reports of “unable to determine”) collapsed into three
19 ordered categories that grouped borderline/marginal with non-borderline/marginal for
20 relevant categories: “may not serve”, “may serve in some settings”, and “may serve in all
21 settings.” We fit an ordinal mixed logistic regression to the supervisory role
22 categorization with AMR and resident year as fixed effects, and random effects for

resident, program, and CCC member, using full Bayesian inference with multi-chain monte carlo sampling via the Stan system²³ and the brms R package.²⁴

The Institutional Review Board at Cincinnati Children's Hospital Medical Center (lead site) and the IRB at each participating program reviewed and approved this study.

Results

Demographic information about programs and CCCs are shown in Table 1. Across 14 participating programs, 68/155 CCC members completed both milestone assignments and supervision categorizations for 451 residents, out of an eligible 852, across all three training years. Of these residents, data were available for 22 in the Fall only, 123 in the Spring only, and 306 in both the Fall and Spring. Ratings were made on the same resident by two different CCC members (e.g., duplicate ratings) for 26 residents (10 in both Spring and Fall, 13 in Fall alone, 3 in Spring alone). These reflect instances where residents are being reviewed by more than one CCC member prior to a full CCC meeting, where consensus decisions will be reached at that meeting. These were not excluded given our intent to determine summative assessment decisions made by individual CCC members rather than to determine a single rating for each resident.

Factor analysis of individual competencies controlling for clustering of residents in raters and sites resulted in a single-factor solution, with cumulative variance of 0.75 and loadings ranging from 0.82 to 0.92. Repeating the analysis separately by PGY year, using fixed rather than random rater effects, and/or excluding the three competencies with 4

1 milestone levels yielded the same result. Factor analysis of the raw ratings (without
2 controlling for clustering) yielded a two-factor solution, with the first factor accounting
3 for 0.70 cumulative variance and the second adding an additional 0.10. The second
4 factor served largely to capture the Prof1 (sense of duty) and Prof4 (help seeking)
5 competencies, suggesting that there may be rater- or program-level differences in how
6 these two competencies are rated but that they do not capture additional resident-related
7 variance after accounting for rater and/or program.

8
9 Most individual competencies had large positive correlations with the AMR (correlation
10 coefficient: 0.84-0.93), with only moderately positive correlations for the professional
11 identity development (Prof1; correlation coefficient: 0.63) and help seeking (Prof4;
12 correlation coefficient: 0.65) competencies (Table 2).

13
14 Table 3 shows the distribution of recommended supervisory role categorizations and
15 AMRs by training year. When combined across training year and time points, the
16 correlation coefficient for the AMR and supervision categorization was 0.56, indicating
17 that 31% of the variance in one is explained by the other (Table 2). For individual
18 competencies, correlation with supervisory role categorization ranged from 0.44-0.56
19 with the exception of Prof1 (professional identity development) and Prof 2 (professional
20 conduct), both 0.34.

21
22 In the ordinal mixed model, higher AMR scores increased the likelihood of being
23 recommended for more supervisory responsibility, both supervising in all settings

1 compared to some settings (OR=1.61, 95% CI 1.45-1.79) and some settings compared to
2 no settings (OR=1.5, 95% CI 1.36-1.67). Resident year was not a significant independent
3 predictor of supervisory role recommendation controlling for AMR. Figure 1 shows each
4 resident's AMR with the color of each dot representing the CCC member recommended
5 supervisory role categorization. The y-axis shows the model's predicted probability that
6 that learner would receive that recommendation. This figure illustrates that the model
7 predicted those who would be granted supervision in all settings well. However, for
8 other recommended supervisory categorizations, the model had a tendency to predict
9 more advanced supervisory levels than CCC members actually recommended for
10 residents. Stated differently, the predictive probability of the AMR on supervisory role
11 categorization was stronger with higher AMRs and less predictive with lower AMRs.

13 **Discussion**

14 This study found moderate to large positive relationships between milestone ratings for
15 individual competencies and AMRs as well as moderately positive relationships between
16 AMRs and supervisory role categorizations.

18 Our factor analysis suggests that we might be able to generate a picture of resident
19 performance by measuring only one or two aspects of performance. If accurate, this
20 could allow the focus of assessment efforts to be narrowed. A competency that
21 synthesizes performance at a high level, such as being considered to be trustworthy or
22 safe, may well be a factor that most or all competencies load on. Future study should
23 explore this further.

1

2 While a single factor solution could simplify assessment efforts, the convergence of
3 within-resident competencies on one factor also raises concerns about CCC member
4 ratings and whether competencies are in fact distinguished or distinguishable. For
5 example, perhaps more rater training on milestones-based assessment is needed among
6 CCC members. Issues of rater training are certainly well-documented.²⁵⁻²⁸ However, it is
7 interesting to note that using narrative descriptions of performance, which is what
8 milestones are, have been shown in previous study to lead to better rater consistency and
9 reliability as well as lead raters to use the range of options available rather than tending to
10 rate high.^{29,30}

11

12 As defined, the 21 competencies for which milestones have been delineated would be
13 expected to measure different things. For example, evidence-based medicine and
14 advocacy are not the same topic. Given this diversity of focus of the 21 competencies,
15 the convergence on a single factor in our analyses was not anticipated. However, this
16 area of milestones research evidence is unclear – are competencies discrete or not? If
17 they are not, should the competencies that residents are assessed on be decreased or
18 should they be redefined to represent a broader range of performance?

19

20 Role of Training Year

21 As Table 3 illustrates, average milestone ratings across residents in the study rose
22 consistently across training year and time of year, with the lowest average milestone
23 rating occurring for PGY1 residents in the Fall (2.66/5) and the highest rating occurring

1 for PGY3 residents in the Spring (4.05/5). This is an anticipated finding. It is not clear
2 based on our study how much these ratings reflect progressive improvement in actual
3 resident performance and how much they reflect assigning progressively higher ratings
4 for residents at progressively higher training levels. However, it should be noted that
5 resident year was not a significant independent predictor of supervisory role
6 recommendation when controlling for AMR in this study. This suggests that CCC
7 members in our study did not use training year to anchor their supervisory role
8 categorizations. This is a positive finding, especially given the exponentially growing
9 international focus on entrustment and supervision as a meaningful assessment
10 framework in medical education.^{3,5,6,11}

11
12 Many PGY1 residents being categorized as able to serve as a supervisor in all settings at
13 the end of the academic year likely reflects their readiness to do what many programs
14 have them do starting in the PGY2 year; namely, supervise other residents. However,
15 categorization to supervise in *all* settings compared to *some* does raise the suspicion that
16 CCC members may not have considered settings were PGY2 residents may not have
17 rotated in yet. If they had considered this, they perhaps would not have chosen all
18 settings. Future research should seek to more clearly capture this.

19 20 Predictive Power of AMR on Supervisory Role Categorization

21 Figure 1 shows that the ordinal mixed model was very good at predicting those who
22 would be granted supervision in all settings. However, for other recommended
23 supervisory role categorizations the model had a tendency to predict a more advanced

1 supervisory role than the CCC member actually recommended for the resident. This may
2 have been impacted by the lower number of responses in the “some settings” category.
3 CCC members may have also based supervisory role decisions on experience rather than
4 demonstrated competency, such as deciding that a resident cannot supervise until rotating
5 through the pediatric intensive care unit regardless of AMR. However, these findings
6 could also reflect a tendency by CCC members to make entrustment decisions that are
7 less lenient when faced with residents whose performance is in the middle ground.

9 Limitations

10 This study has limitations. First, it was conducted in a single specialty, and its results
11 may not be transferable to other specialties. Second, for feasibility purposes, site leads
12 were asked to prospectively recruit a convenience sample of residents that either included
13 all residents or residents representing the full range of performance based on previous
14 performance. However, we did not ask sites to report their sampling strategy and thus
15 can only infer the strategy they took based on their reported data. Third, some
16 participants assigned milestone levels beyond the fourth level when only four levels
17 existed. While this skews data for the three competencies with four milestone levels, a
18 sensitivity analysis in which we excluded these three competencies did not affect the
19 pattern of findings. Fourth, use of the AMR may be an over-generalization of resident
20 assessment. Fifth, 68/155 CCC members completing both milestone ratings and
21 supervisory categorizations could be misinterpreted as a “low response rate.” However,
22 our intent was to collect data based on feasibility at individual sites. While CCC
23 members who did not submit both streams of data may differ from those who did, we

believe our data may still be representative given that it comes from a large number of CCC members across multiple programs. Sixth, there is little validity evidence for the supervisory role categorizations that we studied.¹⁴ Finally, we sought to explore individual CCC member decisions to further inform future study on group CCC member decisions, but it is possible that those group decisions are more accurate and important. Future study should explore the relationship between separate CCC member decisions as well as individual and group decisions.

Conclusion

The moderate to large positive relationships between milestone ratings for individual competencies and AMRs as well as the moderately positive relationships between AMRs and supervisory role categorizations that were observed in this study are encouraging in an assessment era focused on milestones and entrustment. However, this study is exploratory and foundation building and may raise more questions than answers in terms of CCC member ratings and the validity of milestones- and entrustment-based constructs for resident assessment. More validity-focused research is clearly needed given the ubiquity of these constructs in contemporary assessment.

References

1. Nasca TJ, Philibert I, Brigham T, Flynn TC. The Next GME Accreditation System - Rationale and Benefits. *NEJM*. 2012;366:2324-2325.
2. Royal College of Physicians and Surgeons of Canada. The CanMEDS 2015 Framework: Methodology. 2017; <http://www.royalcollege.ca/rcsite/canmeds/about/canmeds-2015-project-methodology-e>. Available at. Accessed May 10, 2017.
3. ten Cate O, Scheele F. Competency-Based Postgraduate Training: Can We Bridge the Gap between Theory and Clinical Practice? *Academic Medicine*. 2007;82(6):542-547.

- 1 **4.** ten Cate O, Snell L, Carraccio C. Medical competence: The interplay between
2 individual ability and the health care environment. *Medical Teacher*. 2010;32:669-
3 675.
- 4 **5.** Chen HC, van den Broek WES, ten Cate O. The Case for Use of Entrustable
5 Professional Activities in Undergraduate Medical Education. *Academic Medicine*.
6 2015;90(4):431-436.
- 7 **6.** Hauer KE, Kohlwes J, Cornett P, et al. Identifying entrustable professional activities
8 in internal medicine training. *Journal of Graduate Medical Education*. 2013;5(1):54-
9 59.
- 10 **7.** Shaughnessy AF, Sparks J, Cohen-Osher M, Goodell KH, Sawin GL, Gravel J.
11 Entrustable professional activities in family medicine. *Journal of Graduate Medical*
12 *Education*. 2013;5(1):112-118.
- 13 **8.** Hicks PJ, Schumacher DJ, Benson B, et al. The Pediatrics Milestones: Conceptual
14 Framework, Guiding Principles, and Approach to Development. *Journal of Graduate*
15 *Medical Education*. 2010;2(3):410-418.
- 16 **9.** American Board of Pediatrics. Entrustable Professional Activities. 2017;
17 <http://www.abp.org/entrustable-professional-activities-epas>. Available at.
18 Accessed May 10, 2017.
- 19 **10.** Carraccio C, Englander R, Holmboe E, Kogan J. Driving Care Quality: Aligning Trainee
20 Assessment and Supervision Through Practical Application of Entrustable
21 Professional Activities, Competencies, and Milestones. *Academic Medicine*.
22 2016;91(2):199-203.
- 23 **11.** Carraccio C, Englander R, Gilhooly J, et al. Building a Framework of Entrustable
24 Professional Activities, Supported by Competencies and Milestones, to Bridge the
25 Educational Continuum. *Academic Medicine*. 2017;92(3):324-330.
- 26 **12.** Warm EJ, Held JD, Hellmann M, et al. Entrusting Observable Practice Activities and
27 Milestones Over the 36 Months of an Internal Medicine Residency. *Academic*
28 *Medicine*. 2016;91(10):1398-1405.
- 29 **13.** Li ST, Tancredi DJ, Schwartz A, et al. Competent for Unsupervised Practice: Use of
30 Pediatric Residency Training Milestones to Assess Readiness. *Academic Medicine*.
31 2017;92(3):385-393.
- 32 **14.** Hicks PJ, Schwartz A. The Story of PMAC: A workplace-based assessment system for
33 the real world. The Philip Dodds Memorial Lecture. Paper presented at:
34 MedBiquitous Annual Conference 2017; Baltimore, MD.
- 35 **15.** Hicks PJ, Margolis M, Poynter SE, et al. The Pediatrics Milestones Assessment Pilot:
36 Development of Workplace-based Assessment Content, Instruments, and Processes.
37 *Academic Medicine*. 2016;91(5):701-709.
- 38 **16.** Turner TL, Bhavaraju VL, Luciw-Dubas UA, et al. Assessment of Pediatric Interns
39 and Sub-Interns on a Subset of Pediatrics Milestones. *Academic Medicine*. 2017;Epub
40 ahead of print
- 41 **17.** Rekman J, Hamstra SJ, Dudek N, Wood T, Seabrook C, Gofton W. A new instrument
42 for assessing resident competence in surgical clinic: the Ottawa clinic assessment
43 tool. *Journal of Surgical Education*. 2016;73(4):575-582.
- 44 **18.** Fabrigar LR, Wegener DT, MacCallum RC, Strahan EJ. Evaluating the use of
45 exploratory factor analysis in psychological research. *Psychological Methods*.
46 1999;4(3):272-299.
- 47 **19.** Horn JL. A rationale and test of the number of factors in factor analysis.
48 *Psychometrika*. 1965;30:179-185.
- 49 **20.** R Core Team. R: A language and environment for statistical computing. 2016;
50 <http://www.R-project.org>. Available at.

- 1 **21.** *nFactors: an R package for parallel analysis and non graphical solutions to the Cattell*
2 *scree test* [computer program]. R package version 2.3.3; 2010.
- 3 **22.** Bernaards CA, Jennrich RI. Gradient Projection Algorithms and Software for
4 Arbitrary Rotation Criteria in Factor Analysis. *Educational and Psychological*
5 *Measurement*. 2005;65:676-696.
- 6 **23.** Team SD. Stan Modeling Language Users Guide and Reference Manual, Version
7 2.14.0. 2016.
- 8 **24.** Bürkner P. brms: An R Package for Bayesian Multilevel Models using Stan. *Journal of*
9 *Statistical Software*. in press.
- 10 **25.** Kogan JR CL, Bernabeo E, Iobst W, Holmboe E. Opening the black box of clinical skills
11 assessment via observation: a conceptual model. *Medical Education*. 2011;45:1048-
12 1060.
- 13 **26.** Holmboe ES WD, Reznick RK, Katsufakis PJ, Leslie KM, Patel V, Ray DD, Nelson EA.
14 Faculty Development in Assessment: The Missing Link in Competency-Based
15 Medical Education. *Academic Medicine*. 2011;86:460-467.
- 16 **27.** Holmboe ES, Hawkins RE, Huot SJ. Effects of training in direct observation of
17 medical residents' clinical competence: A randomized trial. *Ann Intern Med*.
18 2004;140:874-881.
- 19 **28.** Cook DA, Dupras DM, Beckman TJ, Thomas KG, Pankratz VS. Effect of rater training
20 on reliability and accuracy of mini-CEX scores: A randomized, controlled trial. *J Gen*
21 *Intern Med*. 2009;24:74-79.
- 22 **29.** Regehr G, Ginsburg S, Herold J, Hatala R, Eva K, Oulanova O. Using "Standardized
23 Narratives" to Explore New Ways to Represent Faculty Opinions of Resident
24 Performance. *Academic Medicine*. 2012;87(4):419-427.
- 25 **30.** Regehr G, Regehr C, Bogo M, Power R. Can We Build a Better Mousetrap? Improving
26 the Measures of Practice Performance in the Field Practicum. *J Social Work Ed*.
27 2007;43(2):327-344.

28

29

30

1 Table 1: APPD LEARN CCC Study Programs

2

3 Table 2: Correlation Matrix for Competencies and Recommended Supervisory Role
4 Categorization

5

6 Table 3: Recommended Supervisory Role Categorization and Average Milestone Rating,
7 by Resident Year and Time Point

8

9 Figure 1: Predictive Probability of Average Milestone Rating on Recommended
10 Supervisory Role Categorization

11

Figure 1: Predictive Probability of Average Milestone Rating on Recommended Supervisory Role Categorization

Key: Resident Average Milestone Rating (AMR) (x-axis) shown with the color of each dot representing the CCC member recommended supervisory role categorization and the y-axis showing the model's predicted probability that that learner would receive that recommendation.

Table 1: APPD LEARN CCC Study Programs

Program	Boston Combined Residency Program in Pediatrics (Boston Children's Hospital/Boston Medical Center)	Cincinnati Children's Hospital Medical Center	Duke University	Icahn School of Medicine at Mount Sinai	Massachusetts General Hospital	Naval Medical Center San Diego	Phoenix Children's Hospital/Maricopa Medical Center Pediatric Residency Program	St. Christopher's Hospital for Children	University of Arizona	University of California Davis	University of Illinois at Chicago	University of Rochester	University of Texas at Austin	University of Wisconsin
Program Type	Free standing children's hospital with about one-third of time spent at urban safety net hospital with pediatric units within an adult hospital	Free standing children's hospital	Children's hospital within a hospital	Children's hospital within a hospital	Children's hospital within a hospital	Pediatric program in military hospital	Free standing children's hospital	Free standing children's hospital	Children's hospital connected to adult hospital	Children's hospital within a hospital	Children's hospital within a hospital	Free-standing children's hospital with about one-third of time at a community hospital (pediatric floor/units within a hospital)	Free standing children's hospital	Children's hospital connected to adult hospital
Program Size	117 residents	120 residents	45 residents	60 residents	42 residents	22 residents	96 residents	76 residents	48 residents	39 residents	38 residents	44 residents	60 residents	45 residents

Table 2: Correlation Matrix for Competencies and Recommended Supervisory Role Categorization

Key

RSC: Recommended Supervisory Role Categorization

AMR: Average Milestone Rating

	RSC	AMR	PC1	PC2	PC3	PC4	PC5	MK1	PBLI1	PBLI2	PBLI3	PBLI4	Prof1	Prof2	Prof3	Prof4	Prof5	Prof6	SBP1	SBP2	SBP3	ICS1	ICS2
RSC	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
AMR	0.56	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
PC1	0.56	0.93	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
PC2	0.56	0.92	0.9	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
PC3	0.49	0.91	0.88	0.89	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
PC4	0.55	0.93	0.89	0.88	0.86	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
PC5	0.53	0.91	0.86	0.86	0.81	0.88	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
MK1	0.53	0.89	0.85	0.84	0.83	0.84	0.79	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
PBLI1	0.48	0.91	0.84	0.83	0.8	0.82	0.81	0.81	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
PBLI2	0.44	0.84	0.8	0.78	0.76	0.78	0.67	0.8	0.82	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
PBLI3	0.52	0.85	0.78	0.77	0.76	0.77	0.74	0.81	0.75	0.79	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
PBLI4	0.44	0.88	0.8	0.81	0.81	0.8	0.76	0.76	0.83	0.79	0.74	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Prof1	0.34	0.63	0.51	0.51	0.46	0.57	0.69	0.44	0.54	0.28	0.39	0.47	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Prof2	0.46	0.9	0.81	0.8	0.79	0.8	0.81	0.77	0.8	0.73	0.73	0.82	0.62	1	NA	NA	NA	NA	NA	NA	NA	NA	NA
Prof3	0.48	0.92	0.82	0.82	0.79	0.83	0.82	0.78	0.83	0.76	0.75	0.82	0.62	0.88	1	NA	NA	NA	NA	NA	NA	NA	NA
Prof4	0.34	0.65	0.52	0.52	0.48	0.58	0.7	0.45	0.58	0.31	0.39	0.51	0.88	0.61	0.64	1	NA	NA	NA	NA	NA	NA	NA
Prof5	0.5	0.91	0.84	0.84	0.81	0.85	0.84	0.77	0.83	0.74	0.72	0.79	0.61	0.83	0.85	0.66	1	NA	NA	NA	NA	NA	NA
Prof6	0.49	0.91	0.86	0.84	0.85	0.84	0.8	0.83	0.82	0.8	0.79	0.82	0.46	0.8	0.83	0.51	0.81	1	NA	NA	NA	NA	NA
SBP1	0.52	0.92	0.86	0.85	0.84	0.86	0.84	0.84	0.82	0.76	0.78	0.8	0.55	0.81	0.83	0.56	0.83	0.85	1	NA	NA	NA	NA
SBP2	0.51	0.86	0.78	0.8	0.79	0.78	0.75	0.8	0.76	0.74	0.85	0.74	0.47	0.73	0.75	0.46	0.74	0.79	0.81	1	NA	NA	NA
SBP3	0.51	0.89	0.83	0.84	0.84	0.84	0.75	0.82	0.8	0.82	0.8	0.82	0.39	0.79	0.81	0.41	0.78	0.83	0.84	0.79	1	NA	NA
ICS1	0.48	0.91	0.86	0.84	0.82	0.83	0.78	0.82	0.83	0.82	0.78	0.84	0.44	0.81	0.82	0.46	0.82	0.85	0.85	0.77	0.87	1	NA
ICS2	0.5	0.9	0.81	0.8	0.76	0.81	0.84	0.75	0.8	0.69	0.73	0.77	0.68	0.83	0.84	0.7	0.82	0.8	0.83	0.74	0.77	0.82	1

Table 3: Recommended Supervisory Role Categorization and Average Milestone Rating, by Resident Year and Time Point

Year	Time Point	May not Supervise	May Supervise in Some Settings	May Supervise in All Settings	Average Milestone Rating Mean Score (sd)
PGY1	Fall	56 (44%)	42 (33%)	29 (23%)	2.66 (0.60)
PGY1	Spring	3 (2%)	51 (29%)	122 (69%)	2.81 (0.48)
PGY2	Fall	6 (4.4%)	17 (12.3%)	115 (83.3%)	3.39 (0.49)
PGY2	Spring	1 (1%)	8 (4%)	182 (95%)	3.50 (0.39)
PGY3	Fall	0 (0%)	3 (4%)	77 (96%)	3.99 (0.37)
PGY3	Spring	0 (0%)	1 (1%)	72 (99%)	4.05 (0.34)