1-2018

Identifying Gaps in the Performance of Pediatric Trainees Who Receive Marginal/Unsatisfactory Ratings

Su-Ting T. Li
University of California, Davis

Daniel J. Tancredi
University of California, Davis

Alan Schwartz
University of Illinois -College of Medicine

Ann Guillot
University of Vermont - College of Medicine

Ann Burke
Wright State University

See next page for additional authors

Recommended Citation

Li, Su-Ting T. MD, MPH; Tancredi, Daniel J. PhD; Schwartz, Alan PhD; Guillot, Ann MD; Burke, Ann MD; Trimm, R. Franklin MD; Guralnick, Susan MD; Mahan, John D. MD; Gifford, Kimberly A. MD. Identifying Gaps in the Performance of Pediatric Trainees Who Receive Marginal/Unsatisfactory Ratings. Academic Medicine 93(1):p 119-129, January 2018. | DOI: 10.1097/ACM.0000000000001775

Identifying Gaps in the Performance of Pediatric Trainees Who Receive Marginal/Unsatisfactory Ratings

Su-Ting T. Li, MD, MPH, Daniel J. Tancredi, PhD, Alan Schwartz, PhD, Ann Guillot, MD, Ann Burke, MD, R. Franklin Trimm, MD, Susan Guralnick, MD, John D. Mahan, MD, and Kimberly A. Gifford, MD, for the Association of Pediatric Program Directors (APPD) Longitudinal Educational Assessment Research Network (LEARN) Validity of Resident Self-Assessment Group

Abstract

Purpose

To perform a derivation study to determine in which subcompetencies marginal/unsatisfactory pediatric residents had the greatest deficits compared with their satisfactorily performing peers and which subcompetencies best discriminated between marginal/unsatisfactory and satisfactorily performing residents.

Method

Multi-institutional cohort study of all 21 milestones (rated on four or five levels) reported to the Accreditation Council for Graduate Medical Education, and global marginal/unsatisfactory versus satisfactory performance reported to the American Board of Pediatrics. Data were gathered in 2013–2014. For each level of training (postgraduate year [PGY] 1, 2, and 3), mean differences between milestone levels of residents with marginal/unsatisfactory and satisfactory performance adjusted for clustering by program and C-statistics (area under receiver operating characteristic curve) were calculated. A Bonferroni-corrected significance threshold of .0007963 was used to account for multiple comparisons.

Results

Milestone and overall performance evaluations for 1,704 pediatric residents in 41 programs were obtained. For PGY1s, two subcompetencies had almost a one-point difference in milestone levels between marginal/unsatisfactory and satisfactory trainees and outstanding discrimination (≥ 0.90): organize/prioritize (0.93; C-statistic: 0.91) and transfer of care (0.97; C-statistic: 0.90). The largest difference between marginal/unsatisfactory and satisfactory PGY2s was trustworthiness (0.78). The largest differences between marginal/unsatisfactory and satisfactory PGY3s were ethical behavior (1.17), incorporating feedback (1.03), and professionalization (0.96). For PGY2s and PGY3s, no subcompetencies had outstanding discrimination.

Conclusions

Marginal/unsatisfactory pediatric residents had different subcompetency gaps at different training levels. While PGY1s may have global deficits, senior residents may have different performance deficiencies requiring individualized counseling and targeted performance improvement plans.

The public expects competence from physicians.[1–3] The member boards of the American Board of Medical Specialties, together with residency program directors (PDs), certify each individual physician's competence to practice medicine without supervision in that specialty.[4] Pediatric PDs are required to provide an overall assessment of satisfactory, marginal, or unsatisfactory performance to the American Board of Pediatrics (ABP) at the end of each residency year; only physicians who are assessed as satisfactory can become board certified. Little is known about the criteria PDs use to make competence and advancement decisions. Previous reports suggest that the designation of marginal or unsatisfactory performance has been based on general impressions of the trainee.[5] The introduction of educational milestones as a discipline-wide assessment tool by the Accreditation Council for Graduate Medical Education (ACGME) in July 2013 provides a common language for assessment and is a first step in developing a standardized method for examining, across training programs, progression toward becoming independent practitioners.[6,7]

Milestones are observable, competency-based developmental outcomes that learners can demonstrate progressively from the beginning of training through graduation to unsupervised practice. Milestones are organized under six ACGME competency domains: patient care (PC), medical knowledge (MK), interpersonal and communication skills (ICS), practice-based learning and improvement (PBLI), professionalism (Prof), and systems-based practice (SBP).[7] Each specialty worked with the ACGME and their relevant certifying board to create specialty-specific milestones.[7] The pediatric milestones, informed by the literature, describe the stages through which learners progress for each subcompetency. These span the medical education continuum from novice, commensurate with a medical student, to seasoned practicing pediatric expert.[8–11] For most subcompetencies, five milestone levels were defined; however, for some subcompetencies, there was inadequate literature to distinguish between proficiency and mastery; thus, only four milestone levels were defined.[12,13] The milestone level performance of marginal/unsatisfactory (M/U) residents relative to their satisfactorily performing (S) peers is unknown, as are the subcompetencies

Please see the end of this article for information about the authors.

Correspondence should be addressed to Su-Ting T. Li, Department of Pediatrics, University of California, Davis, 2516 Stockton Blvd., Sacramento, CA 95817; telephone: (916) 734-2428; e-mail: sutli@ucdavis.edu.

Acad Med. 2018;93:119–129.
First published online June 20, 2017
doi: 10.1097/ACM.0000000000001775

Supplemental digital content for this article is available at http://links.lww.com/ACADMED/A457.

in which M/U residents have the largest deficits compared with their S peers. By identifying subcompetencies where M/U residents lag, PDs may be able to anticipate areas in which trainees may struggle and thus provide targeted skill development.

The specific aims of our derivation study were threefold: to determine the milestone levels of pediatric residents identified by PDs as either M/U or S, to determine in which subcompetencies M/U pediatric residents had the greatest deficits compared with their S peers, and to determine which subcompetencies best discriminated between M/U pediatric residents and their S peers. We hypothesized that subcompetencies in which M/U pediatric residents had the most difficulty would differ based on level of training.

Method

Study population

We performed a prospective multi-institutional cohort study in academic year 2013–2014, the first year of milestone reporting. PDs were recruited at the 2013 Association of Pediatric Program Directors (APPD) annual spring meeting, as well as through the APPD Longitudinal Educational Assessment Research Network (LEARN)[14] e-mail list.

Data collection

PDs submitted end-of-year (June 2014) data, which corresponded to the ACGME and ABP reporting periods. PDs completed a demographic survey and submitted deidentified resident demographic information, subcompetency milestone levels, and overall PD rating of "satisfactory," "marginal," or "unsatisfactory," as they would submit to the ABP. Program demographics included program size (small [≤ 30 residents], medium [31–60 residents], large [> 60 residents]) and region (Northeast, Midwest, South, West). Resident demographics included gender, medical school (U.S. allopathic medical graduate [USMG-MD], international medical graduate [IMG], U.S. osteopathic medical graduate [USMG-DO]), type of pediatric training (categorical, combined), and level of training (postgraduate year [PGY] 1, PGY2, or PGY3). For the purposes of this study, only data from categorical pediatric residents were used because combined pediatric residents (e.g., medicine–pediatrics) may have a variable number of months of training in pediatrics at different postgraduate years, making it challenging to compare milestone levels across levels of training. PDs reported milestone levels (1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5) for each resident for each of the 21 subcompetencies required by the ACGME. Each milestone level is anchored by behavioral descriptions. Each institution's PD and Clinical Competency Committee determined how residents were assessed to arrive at milestone ratings. In three pediatric subcompetencies (diagnostic/therapeutic decisions [PC4], coordinate care [SBP1], and teamwork [SBP3]), the maximum milestone level is 4 rather than 5. The Institutional Review Boards (IRBs) at the University of California, Davis, and each participating program approved or exempted this study. For the purposes of this deidentified study, IRBs did not require signed consent from each participant given the minimal risk of the research. Some site IRBs required that participants receive a research information sheet with an option to opt out of the study.

Data analysis

We compared the program characteristics of enrolled study programs to unenrolled pediatric programs nationally using the American Medical Association's Fellowship and Residency Electronic Interactive Database Web site.[15] We compared the characteristics of study residents versus estimated unenrolled residents by subtracting study residents from all pediatric residents nationally using ACGME data for categorical pediatric residents from 2013–2014.[16] We compared M/U versus S residents by gender, medical school, and level of training. We performed chi-square analyses to compare study participants versus unenrolled pediatric residents and M/U versus S residents using STATA/SE statistical software, version 12.1 (STATA Corporation, College Station, Texas).

We performed all analyses other than chi-square analyses using SAS statistical software, version 9.4 (SAS Institute, Inc., Cary, North Carolina). We performed a two-sided asymptotic Cochrane–Armitage trend test[17] to determine whether learners were less likely to be assessed as M/U as level of training increased. We calculated mean milestone scores, standard deviations, medians, and interquartile ranges of residents dichotomized into M/U and S for each of the 21 subcompetencies for each level of training. Data for residents with missing level of training (n = 6) were not included in analyses that required level of training. For each level of training, we calculated mean differences between milestone scores of M/U and S residents, adjusted for clustering by program as a fixed effect, to control for confounding effects arising from between-program differences in mean milestone scores and proportion of M/U residents. We did not adjust for resident demographics such as gender or medical school because we felt that any differences in resident overall performance (marginal/unsatisfactory/satisfactory) associated with resident demographics should be captured fully by the resident's milestone scores. A priori, subcompetencies where the adjusted mean difference was ≥ 0.5 milestone level were designated as educationally significant. To determine whether adjustment for resident demographics affected the results, we calculated mean milestone levels adjusted for program, gender, and medical school. For medical school, we combined USMG-MDs and USMG-DOs into one category and IMGs into a second category, as there were no USMG-DOs in the M/U group for some levels of training.[18] We performed an additional regression analysis restricted to S residents to investigate whether S residents in programs reporting M/U residents and programs with no M/U residents had similar milestone scores. To determine whether skew in the data required nonparametric analyses, we rank transformed the data, performed regression analyses on the rank-transformed data, and compared the resulting *P* values with those from our adjusted mean difference analysis.[18] For each of these analyses, we used a Bonferroni adjustment to account for the 63 comparisons such that the *P* value for statistical significance is .0007963 (.05/63) and presented 99.92% (1 − 0.0007963) confidence intervals (CIs) around the effect estimates to reflect the adjusted α level.

To determine how well subcompetency milestone levels discriminated between residents identified as M/U and S, we calculated C-statistics (area under the

Table 1

Demographic Characteristics of 1,704 Pediatric Residents and Unenrolled Pediatric Residents Nationally, From a Multi-Institutional Study of Clinical Skills Progression by Pediatric Milestones, 2013–2014

Resident characteristic	No. (%) of all study pediatric residents			No. (%) of unenrolled pediatric residents nationally[a]	*P* value comparing all study pediatric residents with unenrolled pediatric residents nationally
	Overall N = 1,704	M/U n = 30	S n = 1,674	N = 7,035	
Gender					.41
Female	1,230 (72.2)	17 (56.7)	1,213 (72.5)	4,721 (73.2)	
Male	474 (27.8)	13 (43.3)	461 (27.5)	1,731 (26.8)	
Medical school[b]					< .01
U.S. or Canadian allopathic medical graduate	1,273 (74.8)	20 (66.7)	1,253 (74.9)	4,583 (64.3)	
International medical graduate	276 (16.2)	9 (30.0)	267 (16.0)	1,713 (24.0)	
U.S. osteopathic medical graduate	154 (9.0)	1 (3.3)	153 (9.1)	828 (11.6)	
Level of training[c]					.95
PGY1	570 (33.6)	15 (50.0)	555 (33.3)	2,392 (34.0)	
PGY2	577 (34.0)	9 (30.0)	568 (34.1)	2,381 (33.8)	
PGY3	551 (32.5)	6 (20.0)	545 (32.7)	2,268 (32.2)	

Abbreviations: PGY indicates postgraduate year; M/U, marginal/unsatisfactory; S, satisfactory.
[a]National data from the Accreditation Council for Graduate Medical Education Data Resource Book Academic Year 2013–2014 for United States categorical pediatric residents from 2013–2014 for medical school and sex.[16] The ACGME Data Resource Book Academic Year 2013–2014 had 672 pediatric residents who did not have sex reported. Unenrolled data were estimated by subtracting study data from all data nationally.
[b]1 missing medical school.
[c]6 missing level of training.

receiver operating characteristic curve [AUC]) for each subcompetency for each level of training. We calculated asymptotic 99.92% CIs for the AUC based on Somers D, using Bonferroni adjustment to account for the 63 comparisons. C-statistics can be interpreted as the probability that a randomly selected trainee identified as M/U has a lower milestone score than an S trainee. A C-statistic of 0.5 indicates that a subcompetency is no better than chance at discriminating between M/U and S residents, whereas a C-statistic of 1 indicates that a subcompetency perfectly discriminates between M/U and S residents. C-statistic values of 0.7 to 0.8 indicate acceptable discrimination, values of 0.8 to 0.9 indicate excellent discrimination, and values ≥ 0.9 indicate outstanding discrimination.[19]

Results

A total of 41 pediatric residency programs (20.6%; 41/199 programs in the United States), representing 1,704 unique categorical pediatric residents (570 PGY1s, 577 PGY2s, 551 PGY3s), participated in the study. Compared with nonparticipating programs, study programs were similar in distribution of size (29.3% [12/41] small, 36.6% [15/41] medium, and 34.2% [14/41] large; $P = .36$) and program region (23.4% [10/41] Northeast, 27.7% [12/41] Midwest, 29.8% [10/41] South, 19.2% [9/41] West; $P = .31$). The demographic distribution of participants reflects those of all United States pediatric residents in terms of postgraduate year and gender (Table 1).[16] Our study had more USMG-MDs than the unenrolled group ($P < .01$; 74.8% vs. 64.3%).

Residents were more likely to be reported as M/U earlier in training (PGY1: 2.6% [15/570]; PGY2: 1.6% [9/577]; PGY3: 1.1% [6/551]; $P = .0495$). There was no difference in gender ($P = .06$) or medical school ($P = .08$) between M/U and S resident groups.

Mean milestone levels of M/U and S residents

Table 2 and Figure 1 show how M/U residents performed relative to their S peers. S PGY1s' end-of-year mean milestone levels were ~2.5 to 3 for each subcompetency, with a ~0.5-milestone-level increase with each additional year of training. Figure 1's radar graph shows a clear difference between performance of M/U and S PGY1s, with M/U PGY1s performing 0.5 to 1 milestone level below their S peers in all subcompetency areas. Mean milestone levels of M/U PGY2s were similar to S PGY1s' mean levels in trustworthiness (Prof5). Mean milestone levels of M/U PGY3s were lower than S PGY1s' mean levels for ethical behavior (Prof3).

Table 2 and Figure 2 show that each level of training had different subcompetencies where M/U residents' adjusted mean milestone level performance differed from their S peers. Results were similar when mean differences were adjusted for program, gender, and medical school (data not shown). Results from our nonparametric (rank transformed) analysis were similar to results from our parametric (adjusted mean difference) analysis (data not shown), with the exception of three items which were no longer statistically significant: PGY1–advocacy (SBP2), PGY2–evidence-based pediatrics (MK), and PGY3–diagnostic/therapeutic decisions (PC4). For PGY1s, M/U residents had globally lower milestone levels than their S peers for both the parametric and nonparametric

Table 2

End-of-Year Adjusted Mean Difference in Milestone Levels Between S and M/U Pediatric Residents by Level of Training, 2013–2014[a]

Subcompetency	PGY1			PGY2			PGY3		
	M/U n = 15: mean (SD) [median, 25%–75%]	S n = 555: mean (SD) [median, 25%–75%]	Adjusted mean difference (99.92% CI)	M/U n = 9: mean (SD) [median, 25%–75%]	S n = 568: mean (SD) [median, 25%–75%]	Adjusted mean difference (99.92% CI)	M/U n = 6: mean (SD) [median, 25%–75%]	S n = 545: mean (SD) [median, 25%–75%]	Adjusted mean difference (99.92% CI)
PC1: Gather essential and accurate information about the patient	1.80 (0.59) [2, 1–2.5]	2.76 (0.60) [2.5, 2.5–3]	0.76 (0.39 to 1.12)[c]	3.17 (0.71) [3, 2.5–4]	3.43 (0.59) [3.5, 3–4]	0.57 (0.23 to 0.91)[c]	3.50 (0.94) [4, 2.5–4]	3.95 (0.52) [4, 3.5–4.5]	0.53 (–0.36 to 1.41)
PC2: Organize and prioritize responsibilities to provide patient care that is safe, effective, and efficient	1.63 (0.52) [1.5, 1.5–2]	2.73 (0.61) [2.5, 2.5–3]	0.93 (0.54 to 1.32)[c]	3.00 (0.61) [2.5, 2.5–3.5]	3.41 (0.63) [3.5, 3–4]	0.60 (0.29 to 0.91)[c]	3.42 (1.02) [3.5, 2.5–4.5]	3.95 (0.53) [4, 3.5–4.5]	0.63 (–0.20 to 1.45)
PC3: Provide transfer of care that ensures seamless transitions	1.70 (0.56) [1.5, 1–2]	2.82 (0.66) [3, 2.5–3]	0.97 (0.60 to 1.34)[c]	2.93 (0.45) [3, 2.5–3.5]	3.43 (0.60) [3.5, 3–4]	0.51 (0.16 to 0.86)[c]	3.25 (1.33) [3.5, 2.5–4.5]	3.93 (0.53) [4, 3.5–4]	0.99 (–0.34 to 2.32)
PC4[b]: Make informed diagnostic and therapeutic decisions that result in optimal clinical judgment	1.70 (0.49) [1.5, 1.5–2]	2.67 (0.55) [2.5, 2.5–3]	0.82 (0.54 to 1.10)[c]	2.89 (0.60) [2.5, 2.5–3.5]	3.30 (0.55) [3.5, 3–3.5]	0.62 (0.08 to 1.16)[c]	3.25 (0.61) [3.5, 2.5–3.5]	3.79 (0.48) [4, 3.5–4]	0.57 (0.06 to 1.08)
PC5: Develop and carry out management plans	1.70 (0.59) [1.5, 1.5–2]	2.76 (0.58) [3, 2–3]	0.88 (0.49 to 1.28)[c]	3.11 (0.60) [3, 2.5–3.5]	3.41 (0.59) [3.5, 3–4]	0.57 (0.12 to 1.02)[c]	3.42 (0.97) [4, 2.5–4]	3.94 (0.52) [4, 3.5–4.5]	0.71 (–0.13 to 1.56)
MK: Locate, appraise, and assimilate evidence from scientific studies related to their patients' health problems	1.77 (0.60) [2, 1.5–2]	2.72 (0.63) [2.5, 2.5–3]	0.73 (0.30 to 1.16)[c]	3.11 (0.60) [3, 2.5–3.5]	3.32 (0.61) [3.5, 3–3.5]	0.52 (0.01 to 1.03)	3.83 (0.93) [4, 2.5–4.5]	3.82 (0.55) [4, 3.5–4]	0.23 (–0.60 to 1.06)
PBLI1: Identify strengths, deficiencies, and limits in one's knowledge and expertise	1.79 (0.55) [1.5, 1.5–2]	2.85 (0.63) [3, 2.5–3]	0.85 (0.52 to 1.18)[c]	2.94 (0.85) [3, 2–3.5]	3.37 (0.61) [3.5, 3–4]	0.66 (0.13 to 1.18)[c]	3.42 (0.97) [4, 2.5–4]	3.84 (0.53) [4, 3.5–45]	0.61 (–0.12 to 1.35)
PBLI2: Identify and perform learning activities to guide personal and professional development	1.73 (0.56) [1.5, 1.5–2]	2.71 (0.70) [2.5, 2–3]	0.79 (0.50 to 1.09)[c]	2.94 (0.81) [2.5, 2.5–3]	3.22 (0.65) [3, 3–3.5]	0.63 (0.00 to 1.27)	3.50 (0.89) [3.5, 2.5–4.5]	3.79 (0.58) [4, 3.5–4]	0.52 (–0.15 to 1.20)
PBLI3: Systematically analyze practice using quality improvement methods, and implement changes with the goal of practice improvement	1.77 (0.62) [2, 1–2]	2.50 (0.57) [2.5, 2–3]	0.60 (0.26 to 0.94)[c]	2.57 (0.73) [2.5, 2–3]	3.11 (0.59) [3, 2.5–3.5]	0.63 (0.07 to 1.20)[c]	2.50 (1.32) [3, 1–3.5]	3.64 (0.54) [3.5, 3–4]	0.85 (–0.91 to 2.61)
PBLI4: Incorporate formative evaluation feedback into daily practice	1.90 (0.74) [2, 1.5–2.5]	2.94 (0.65) [3, 2.5–3.5]	0.80 (0.29 to 1.32)[c]	3.17 (0.83) [3, 2.5–4]	3.46 (0.65) [3.5, 3–4]	0.48 (–0.09 to 1.04)	3.08 (1.07) [3, 2–4]	3.91 (0.54) [4, 3.5–4]	1.03 (0.25 to 1.81)[c]
ICS1: Communicate effectively with patients, families, and the public, as appropriate, across a broad range of socioeconomic and cultural backgrounds	2.10 (0.57) [2, 1.5–2.5]	2.99 (0.67) [3, 2.5–3.5]	0.66 (0.25 to 1.06)[c]	3.50 (0.97) [3.5, 3–4]	3.53 (0.66) [3.5, 3–4]	0.23 (–0.25 to 0.71)	3.67 (0.88) [3, 3–4.5]	3.98 (0.60) [4, 3.5–4.5]	0.41 (–0.23 to 1.05)
ICS2: Demonstrate the insight into emotion and human response to emotion that allows one to appropriately develop and manage human interactions	1.93 (0.76) [1.5, 1.5–2.5]	2.90 (0.66) [3, 2.5–3.5]	0.75 (0.26 to 1.25)[c]	3.00 (0.75) [3.5, 2.5–3.5]	3.47 (0.68) [3.5, 3–4]	0.68 (0.11 to 1.24)[c]	3.10 (0.89) [3, 2.5–4]	3.92 (0.58) [4, 3.5–4.5]	0.81 (0.02 to 1.61)[c]

(Table continues)

Table 2
(Continued)

Subcompetency	PGY1 M/U n = 15: mean (SD) [median, 25%–75%]	PGY1 S n = 555: mean (SD) [median, 25%–75%]	PGY1 Adjusted mean difference (99.92% CI)	PGY2 M/U n = 9: mean (SD) [median, 25%–75%]	PGY2 S n = 568: mean (SD) [median, 25%–75%]	PGY2 Adjusted mean difference (99.92% CI)	PGY3 M/U n = 6: mean (SD) [median, 25%–75%]	PGY3 S n = 545: mean (SD) [median, 25%–75%]	PGY3 Adjusted mean difference (99.92% CI)
Prof1: Humanism, compassion, integrity, and respect for others; 3.10 (0.63) based on the characteristics of an empathetic practitioner	2.30 (0.73) [2, 2–3]	3.12 (0.61) [3, 3–3.5]	0.70 (0.20 to 1.19)[c]	3.50 (0.56) [3.5, 3–4]	3.60 (0.62) [3.5, 3–4]	0.38 (–0.24 to 0.99)	3.75 (0.61) [3.5, 3.5–4.5]	3.97 (0.54) [4, 3.5–4.5]	0.42 (–0.10 to 0.94)
Prof2: Professionalization: A sense of duty and accountability to patients, society, and the profession	2.23 (0.80) [2, 2–3]	3.10 (0.63) [3, 3–3.5]	0.76 (0.17 to 1.35)[c]	3.33 (0.71) [3, 3–4]	3.59 (0.63) [3.5, 3–4]	0.54 (–0.01 to 1.08)	3.25 (1.08) [3.5, 2–4]	4.00 (0.55) [4, 3.5–4.5]	0.96 (0.14 to 1.79)[c]
Prof3: Professional conduct: High standards of ethical behavior which includes maintaining appropriate professional boundaries	2.30 (0.82) [2, 2–3]	3.05 (0.66) [3, 2.5–3.5]	0.64 (0.11 to 1.16)[c]	3.31 (0.53) [3, 3–4]	3.47 (0.60) [3.5, 3–4]	0.36 (–0.07 to 0.80)	2.75 (1.19) [2, 2–3.5]	3.97 (0.57) [4, 3.5–4.5]	1.17 (0.26 to 2.08)[c]
Prof4: Self-awareness of one's own knowledge, skill, and emotional limitations that lead to appropriate help-seeking behaviors	1.93 (0.76) [2, 1.5–2]	3.04 (0.70) [3, 2.5–3.5]	0.96 (0.46 to 1.47)[c]	3.13 (0.92) [3, 2–4]	3.50 (0.66) [3.5, 3–4]	0.52 (–0.07 to 1.11)	3.50 (1.41) [3.5, 2.5–5]	3.95 (0.54) [4, 3.5–4.5]	0.71 (–0.48 to 1.90)
Prof5: Trustworthiness that makes colleagues feel secure when one is responsible for the care of patients	1.97 (0.77) [2, 1.5–2.5]	3.00 (0.65) [3, 2.5–3.5]	0.83 (0.29 to 1.38)[c]	2.94 (0.73) [2.5, 2.5–3.5]	3.55 (0.65) [3.5, 3–4]	0.78 (0.36 to 1.20)[c]	3.33 (1.29) [3.5, 3–4.5]	4.00 (0.56) [4, 3.5–4.5]	0.88 (–0.44 to 2.19)
Prof6: The capacity to accept that ambiguity is part of clinical medicine and to recognize the need for and to utilize appropriate resources in dealing with uncertainty	1.83 (0.62) [1.5, 1.5–2]	2.79 (0.62) [3, 2.5–3]	0.77 (0.33 to 1.22)[c]	3.06 (0.81) [3, 2.5–3.5]	3.36 (0.64) [3.5, 3–4]	0.56 (–0.01 to 1.13)	3.50 (0.95) [3.5, 3–4.5]	3.88 (0.54) [4, 3.5–4]	0.50 (–0.31 to 1.30)
SBP1[b]: Coordinate patient care within the health care system relevant to their clinical specialty	1.80 (0.56) [2, 1.5–2]	2.71 (0.59) [2.5, 2.5–3]	0.80 (0.36 to 1.23)[c]	3.00 (0.53) [3, 3–3.5]	3.31 (0.57) [3.5, 3–3.5]	0.62 (0.35 to 0.89)[c]	3.42 (0.58) [3.5, 3–4]	3.76 (0.49) [4, 3.5–4.5]	0.33 (–0.02 to 0.67)
SBP2: Advocate for quality patient care and optimal patient care systems	1.81 (0.80) [1.5, 1–2]	2.57 (0.67) [2.5, 2–3]	0.65 (0.11 to 1.19)	2.86 (0.85) [2.5, 2–4]	3.15 (0.65) [3, 2.5–3.5]	0.66 (0.20 to 1.12)[c]	2.67 (0.76) [2.5, 2–3.5]	3.70 (0.60) [4, 3–4]	0.58 (–0.14 to 1.31)
SBP3[b]: Work in interprofessional teams to enhance patient safety and improve patient care quality	1.83 (0.79) [1.5, 1–2.5]	2.85 (0.63) [3, 2.5–3]	0.83 (0.35 to 1.30)[c]	3.06 (0.88) [3, 2.5–3.5]	3.38 (0.59) [3.5, 3–4]	0.66 (0.18 to 1.14)[c]	3.50 (0.77) [3.5, 3–4]	3.81 (0.51) [4, 3.5–4]	0.39 (–0.17 to 0.94)

Abbreviations: M/U indicates marginal/unsatisfactory; S, satisfactory; PGY, postgraduate year; PC, patient care; MK, medical knowledge; PBLI, practice-based learning and improvement; ICS, interpersonal and communication skills; Prof, professionalism; SBP, systems-based practice.
[a]Mean differences are adjusted for residency program. A Bonferroni-corrected significance threshold of .0007963 (.05/63) was used to account for the 63 comparisons.
[b]These subcompetencies have only four milestone levels.
[c]These differences are also statistically significant in nonparametric (regression analysis of rank-transformed data) analyses. A Bonferroni-corrected significance threshold of .0007963 (.05/63) was used to account for the 63 comparisons.

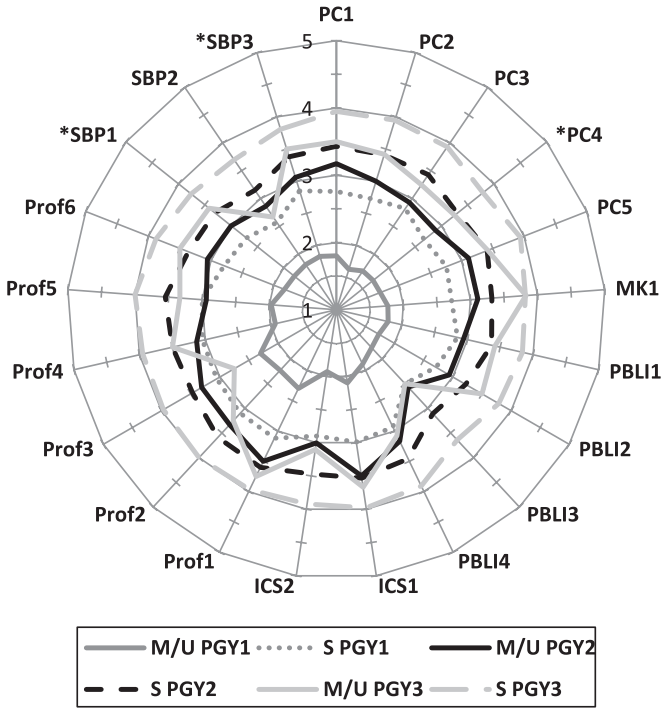

Figure 1 End-of-year subcompetency mean milestone levels for M/U and S pediatric residents by level of training, from a multi-institutional study of clinical skills progression by pediatric milestones, 2013–2014. The radar chart graphically displays mean milestone levels by level of training for M/U and S pediatric residents for each of the 21 pediatric subcompetencies along an axis that begins with 1 in the center of the chart and ends with 5 (highest milestone level) on the outer ring. (For a full listing of the subcompetencies see Table 1 or Appendix 1.) Pediatric residents with satisfactory performance have higher milestone levels with each additional level of training. M/U PGY1s score 0.5–1 milestone level below their S peers in all subcompetency areas. Mean milestone levels of M/U PGY2s were similar to S PGY1s' mean levels in trustworthiness (Prof5). Mean milestone levels of M/U PGY3s were lower than S PGY1s' mean levels for ethical behavior (Prof3). Abbreviations: M/U indicates marginal/unsatisfactory; S, satisfactory; PGY, postgraduate year; PC, patient care; MK, medical knowledge; PBLI, practice-based learning and improvement; ICS, interpersonal and communication skills; Prof, professionalism; SBP, systems-based practice.
*These subcompetencies have only four milestone levels.

analyses for all subcompetencies except advocacy (SBP2). Differences in mean milestone levels ranged from 0.60 to 0.97. Subcompetencies with almost one-full-milestone-level difference between M/U and S PGY1s included organize/prioritize (PC2: 0.93; 99.92% CI: 0.54–1.32), transfer of care (PC3: 0.97; 99.92% CI: 0.60–1.34), and help-seeking (Prof4: 0.96; 99.92% CI: 0.46–1.47). For PGY2s, the largest adjusted mean difference in milestone levels was in trustworthiness (Prof5: 0.78; 99.92% CI: 0.36–1.20). For PGY3s, the largest adjusted mean milestone level differences were ethical behavior (Prof3: 1.17; 99.92% CI: 0.26–2.08), incorporating feedback (PBLI4: 1.03; 99.92% CI: 0.25–1.81), and professionalization (0.96; 99.92% CI: 0.14–1.79). We found no significant difference in milestone levels for S residents in programs reporting M/U residents and programs with no M/U residents (data not shown).

Discrimination between M/U and S residents

Figure 3 and Appendix 1 show that for different levels of training, different subcompetencies discriminated between M/U and S learners. For PGY1s, although all subcompetencies acceptably discriminated between M/U and S learners (C-statistic ≥ 0.7), three subcompetencies had outstanding discrimination (C-statistic ≥ 0.9): organize/prioritize (PC2: C-statistic: 0.91; 99.92% CI: 0.78–0.1.00), transfer of care (PC3: 0.90; 99.92% CI: 0.84–0.96), and diagnostic/therapeutic decisions (PC4: 0.90; 99.92% CI: 0.83–0.96) (see also Supplemental Digital Appendix 1 at http://links.lww.com/ACADMED/A457). For PGY2s, no subcompetency discriminated between M/U and S learners. Trustworthiness (Prof5), which was found to have the largest mean milestone level difference between M/U and S PGY2s, was not significantly discriminatory (C-statistic: 0.73; 99.92% CI: 0.42–1.00). For PGY3s, two subcompetencies had excellent discrimination (C-statistic ≥ 0.8): quality improvement (PBLI3: 0.82; 99.92% CI: 0.50–1.00) and advocacy (SBP2: 0.87; 99.92% CI: 0.56–1.00). The three subcompetencies that had the largest mean milestone-level difference between M/U and S PGY3s were not significantly discriminatory: incorporating feedback (PBLI4: 0.73; 99.92% CI: 0.28–1.00), professionalization (Prof2: 0.70; 99.92% CI: 0.28–1.00), and ethical conduct (Prof3: 0.79; 99.92% CI: 0.21–1.00).

Discussion

Our study is the first, to our knowledge, to use milestone ratings to describe the performance of a nationally representative sample of M/U and S pediatric residents. We found fewer M/U residents as training level progressed. Overall, mean milestone levels for M/U residents were lower than for S residents. M/U residents had different distributions of subcompetency gaps compared with S residents at different levels of training.

PGY1s who were identified as M/U received lower adjusted mean milestone scores than their S peers across all subcompetencies. We identified two subcompetencies that identified residents in need of remediation early in residency because they showed almost a one-point milestone-level difference between M/U and S PGY1s and had outstanding discrimination: organize/prioritize (PC2) and transfer of care (PC3). These findings are consistent with earlier studies suggesting that typical patient care deficits are identifiable early and may be targeted for remediation earlier in training.[20–25]

For PGY2s, the differences in adjusted mean milestone levels were smaller than for PGY1s. Trustworthiness was the subcompetency with the largest adjusted mean milestone difference between M/U and S PGY2s, with M/U PGY2s performing similarly to S PGY1s. In pediatrics, PGY2s begin to have a supervisory role, which may highlight the importance of "trustworthiness that makes colleagues feel secure when one is responsible for the care of patients"[26] in entrustment decisions when allowing a

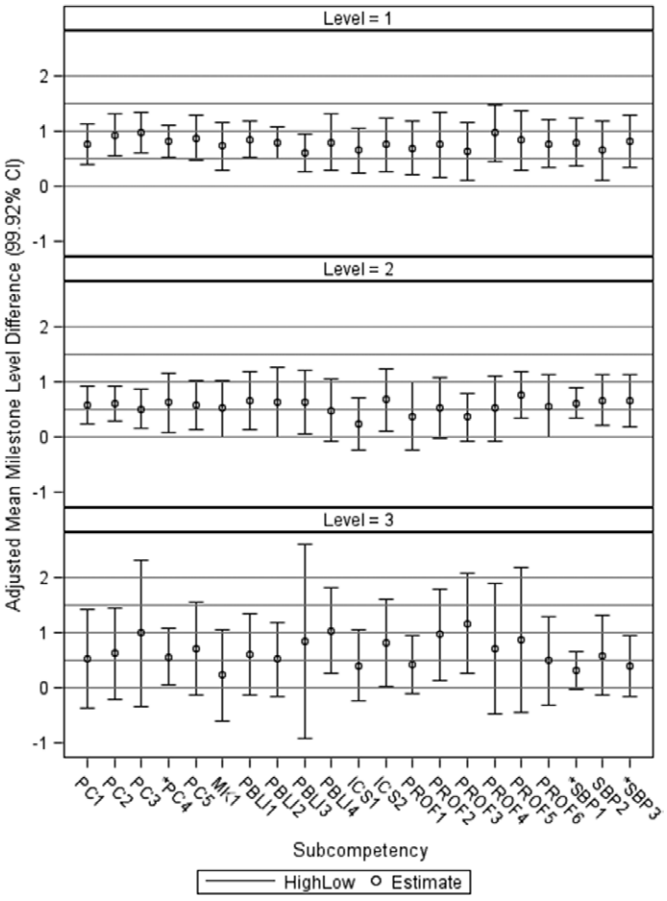

Figure 2 Difference between end-of-year subcompetency mean milestone scores of marginal/unsatisfactory pediatric residents and satisfactorily performing peers by level of training adjusted for program, from a multi-institutional study of clinical skills progression by pediatric milestones, 2013–2014. Mean differences in milestone levels between marginal/unsatisfactory pediatric residents and satisfactorily performing peers are adjusted for clustering by residency program and corrected for the 63 comparisons using a Bonferroni-corrected significance level of .0007963 (.05/63). Adjusted mean differences and 99.92% CIs (1 – 0.0007963) are reported. Results from the nonparametric analysis (regression analysis of rank-transformed data adjusted for clustering by program and with Bonferroni correction) are similar to parametric results reported, with the exception of three items which are no longer statistically significant: PGY1–advocacy (SBP2), PGY2–evidence-based pediatrics (MK), and PGY3–diagnostic/therapeutic decisions (PC4). Abbreviations: PGY indicates postgraduate year; PC, patient care; MK, medical knowledge; PBLI, practice-based learning and improvement; ICS, interpersonal and communication skills; Prof, professionalism; SBP, systems-based practice; CI, confidence interval.
*These subcompetencies have only four milestone levels.

resident to supervise.[27,28] Trustworthiness includes colleagues' perception of the physician's knowledge, skills, and abilities as well as discernment of limitations, conscientiousness, and truthfulness, and may represent how the overall development of entrustment for S PGY2s differs from that for M/U PGY2s.[27,29–31] However, in contrast to M/U PGY1s, who scored globally lower than their S peers, with most subcompetencies having outstanding discrimination (C-statistic ≥ 0.90) or excellent discrimination (C-statistic ≥ 0.80) between M/U and S PGY1s, no subcompetency discriminated between M/U and S PGY2s. This finding points to the importance of PDs individualizing performance improvement plans for M/U PGY2s based on their own personal performance lapses.

We found that mean milestone scores of M/U PGY3s were lower than S PGY1s for ethical behavior (Prof3). In addition, compared with S PGY3s, M/U PGY3s had the largest adjusted mean milestone differences in ethical behavior (Prof3), incorporating feedback (PBLI4), and professionalization (0.96). However, these subcompetencies did not discriminate between M/U and S PGY3s. These findings highlight the heterogeneity of M/U PGY3 performance; some, but not all M/U PGY3s had clear deficiencies in ethical behavior, incorporating feedback, or professionalization. Prior studies have also found that professionalism deficits tend to be identified later in training.[22,32] Difficulties with professionalism or incorporating feedback may be better identified later in training because of increased observations of the trainee. Alternatively, these findings may suggest that successful performance in subcompetencies such as ethical behavior and incorporating feedback may be areas that PDs feel are important in making global M/U or S end-of-year assessments of resident performance. PDs may be making decisions about graduation in part based on professionalism characteristics demonstrated by medical students and residents that may be associated with future medical board disciplinary action, such as lack of professionalization, ethical behavior, and incorporating feedback.[5,33–35]

Our study has several limitations. First, PDs identified a small number of M/U residents. Fewer M/U residents were identified later in training, leading to wider CIs for PGY2 and PGY3 results, and this may underpower our study to detect significant differences between M/U and S resident (Type II errors). It is possible that fewer M/U residents were identified because M/U residents identified earlier in training are remediated or dismissed. Alternatively, PDs may be more reluctant to label residents further in training as M/U because the stakes are higher, with residents' approval to sit for the board certification exam dependent on PD attestation of satisfactory performance and ability to practice without supervision. Second, our study was performed when milestones were new to programs, faculty, and Clinical Competency Committees. Inexperience may have resulted in higher ratings of high-performing residents and lower ratings of low-performing residents, which may lead to overestimation of differences between milestone levels for M/U versus S residents. Conversely, because data were collected early after implementation of milestones, there were no national expectations that may have otherwise biased scoring by the evaluators. Adjusted mean milestone

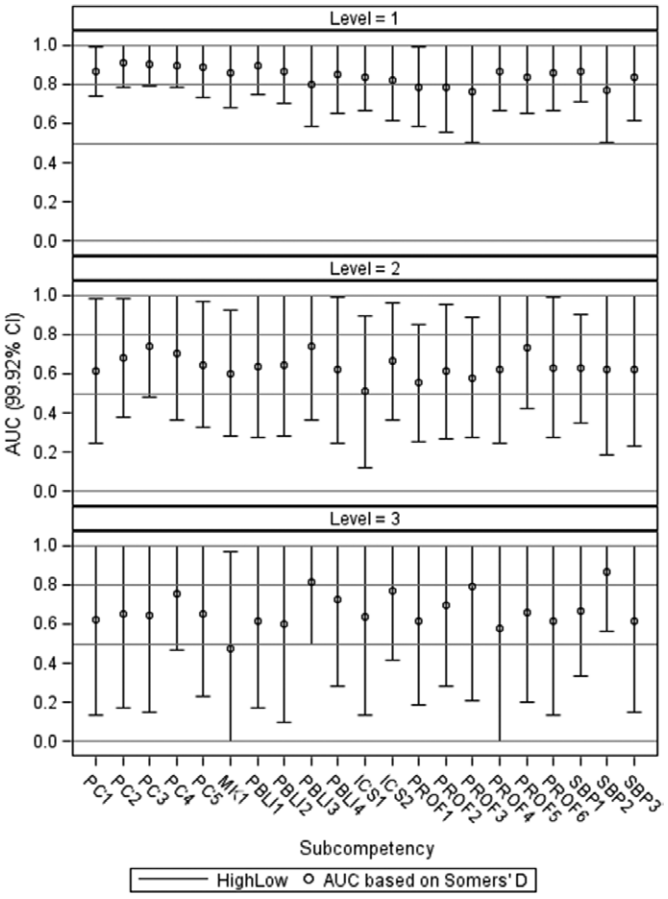

Figure 3 Discrimination between M/U and S pediatric residents for different subcompetencies by level of training, from a multi-institutional study of clinical skills progression by pediatric milestones, 2013–2014. The C-statistic, or AUC, was calculated to determine the ability of different subcompetencies to discriminate between M/U and S pediatric residents, and 99.92% CIs based on Somers D were calculated using a Bonferroni-corrected significance level of .0007963 (.05/63) to correct for the 63 comparisons. C-statistics can be interpreted as the probability that a randomly selected trainee identified as M/U has a lower milestone score than an S trainee. A C-statistic of 0.5 indicates that a subcompetency is no better than chance at discriminating between M/U and S residents, whereas a C-statistic of 1 indicates that a subcompetency perfectly discriminates between M/U and S residents. C-statistic values of 0.7–0.8 indicate acceptable discrimination, values of 0.8–0.9 indicate excellent discrimination, and values ≥ 0.9 indicate outstanding discrimination. Abbreviations: AUC indicates area under receiver operating characteristic curve; M/U, marginal/unsatisfactory; S, satisfactory; CI, confidence interval; PC, patient care; MK, medical knowledge; PBLI, practice-based learning and improvement; ICS, interpersonal and communication skills; Prof, professionalism; SBP, systems-based practice.

differences and C-statistics for PGY3s were variable, suggesting that residents were not given the same milestone level across all subcompetencies. Third, our study had multiple comparisons, which may inadvertently lead to detecting differences where no true differences exist (Type I errors); however, we tried to minimize Type I errors by using a Bonferroni-corrected significance threshold of .0007963 (.05/63) to correct for multiple comparisons. Fourth, our study population had more USMG-MDs compared with all U.S. pediatric residents. It is possible that USMG-DOs and IMGs may have different deficiencies. Future studies in a more representative population of pediatric residents are needed to validate the findings in this derivation study. Fifth, programs may use different criteria to define M/U residents (milestone scores, narrative comments, etc.) and may have different thresholds for determining M/U (e.g., marginal/unsatisfactory residents perform poorly in all milestones, majority of milestones, or specific milestones). Sixth, our study was limited to pediatric residents; it is possible that residents in other specialties have different subcompetency gaps. Future studies of residents of other specialties are needed to verify whether our findings are generalizable beyond pediatrics. Finally, our study was limited to a single year. A longitudinal study which follows residents across training years may help us better understand the progression of M/U residents.

We found that M/U pediatric residents had different subcompetency gaps at different levels of training. While PGY1s have global deficits, senior residents may have different performance deficiencies that require individualized counseling and targeted performance improvement plans. Similar research in a larger validation cohort of pediatric residents, and in other specialties, that uses the subcompetencies found to identify M/U learners in this derivation study is needed to verify or contradict our findings and determine whether our findings are generalizable beyond pediatrics.

Acknowledgments: The authors wish to thank Robin Young, MS, from APPD LEARN (Association of Pediatric Program Directors, Longitudinal Educational Assessment Research Network), for assistance with this project. The authors wish to thank the participating residents and program directors.

The authors wish to thank the site investigators for the APPD LEARN Validity of Resident Self-Assessment Group: Dennis Basila, MD, and Katherine Dougherty, MD (Albany Medical Center); Richard S. Robus, MD (Blank Children's Hospital); Christa Matrone, MD, and Daniel J. Schumacher, MD, MEd (Boston Combined Residency Program); Kimberly A. Gifford, MD (Children's Hospital at Dartmouth-Hitchcock Pediatric Residency); Stephanie B. Dewar, MD, Dena Hofkosh, MD, MEd, and Rhett Lieberman, MD, MPH (Children's Hospital of Pittsburgh of UPMC); Sue E. Poynter, MD, MEd (Cincinnati Children's/University of Cincinnati); Dawn Tuell, MD (East Tennessee State University); Janara J. Huff, MD, and Marielisa Rincon, MD (Erlanger [UT-COM Chattanooga]); Kerry K. Sease, MD, MPH (Greenville Health System/University of South Carolina–Greenville); Sylvia H. Yeh, MD (Harbor-UCLA Medical Center); Michael P. McKenna, MD, and Laura Price, MD (Indiana University); Kathleen M. Donnelly, MD, and Meredith L. Carter, MD (Inova Children's Hospital); David I. Rappaport, MD (Jefferson Medical College/AI DuPont Hospital for Children); Robert R. Wittler, MD (Kansas University School of Medicine–Wichita); Ariel Frey-Vogel, MD, MAT, and Shannon E. Scott-Vernaglia, MD (MassGeneral Hospital for Children); Kate Perkins, MD, PhD, and Savanna Carson, MS (Mattel Children's Hospital UCLA); John D. Mahan, MD (Nationwide Hospital/Ohio State University); Matthew J. Kapklein, MD, MPH (New York Medical College at Westchester

Medical Center); Sharon M. Unti, MD (Ann & Robert H. Lurie Children's Hospital of Chicago/Northwestern University); Cynthia L. Ferrell, MD, MSEd (Oregon Health & Science University); Alston E. Dunbar, MD, MBA (Our Lady of the Lake Regional Medical Center); Vasudha L. Bhavaraju, MD, and Grace L. Caputo, MD, MPH (Phoenix Children's Hospital/Maricopa Medical Center); Jane E. Kramer, MD (Rush University); Michelle Miner, MD (Southern Illinois University); Sharon Calaman, MD, Mario Cruz, MD, Nancy D. Spector, MD (St Christopher's Hospital for Children/Drexel University College of Medicine); Stephen R. Barone, MD (Steven and Alexandra Cohen Children's Medical Center of New York); Tammy Camp, MD (Texas Tech University Health Science Center–Lubbock); Sean P. Elliott, MD, and Hillary A. Franke, MD, MS (University of Arizona); Su-Ting T. Li, MD, MPH (University of California, Davis); Heather Fagan, MD, and Nicola Orlov, MD (University of Chicago); Tai M. Lockspeiser, MD, MHPE (University of Colorado); Nicole Paradise Black, MD, MEd (University of Florida–Gainesville); Jose J. Zayas, DO (University of Florida, College of Medicine–Jacksonville); Jennifer Di Rocco, DO, and Shilpa J. Patel, MD (University of Hawai'i); Amanda D. Osta, MD (University of Illinois–Chicago); Lisa Gilmer, MD (University of Kansas–Kansas City); Sara Multerer, MD, and Kimberly A. Boland, MD (University of Louisville); Rosina A. Connelly, MD, MPH, and R. Franklin Trimm, MD (University of South Alabama), Diane Straub, MD, and Meaghan Reyes (University of South Florida); Sandra R. Arnold, MD (University of Tennessee–Memphis); Deirdre A. Caplin, PhD (University of Utah); Ann P. Guillot, MD, and Jerry G. Larrabee, MD (University of Vermont); John G. Frohna, MD, MPH (University of Wisconsin); Renuka Verma, MD (Unterberg Children's Hospital at Monmouth Medical Center); Jill Leavens-Maurer, MD (Winthrop University Hospital); and Ann E. Burke, MD (Wright State).

Funding/Support: None reported.

Other disclosures: None reported.

Ethical approval: The Institutional Review Board (IRB) at the University of California, Davis (IRB ID: 444607-1; date of IRB approval: 4/23/2013) and the IRBs from each site approved this study.

Previous presentations: This work was presented as a platform presentation at the Pediatric Academic Societies' Annual National Conference in San Diego, California; May 25–28, 2015.

S.T. Li is associate professor, vice chair of education, and pediatric program director, Department of Pediatrics, University of California, Davis, Sacramento, California.

D.J. Tancredi is associate professor, Department of Pediatrics and Center for Healthcare Policy and Research, University of California, Davis, Sacramento, California.

A. Schwartz is Michael Reese Endowed Professor of Medical Education, associate head, and director of research, Department of Medical Education, and research professor, Department of Pediatrics, University of Illinois College of Medicine, Chicago, Illinois.

A. Guillot is professor, Department of Pediatrics, University of Vermont College of Medicine, Burlington, Vermont.

A. Burke is professor and pediatric program director, Department of Pediatrics, Wright State University Boonshoft School of Medicine, Dayton, Ohio.

R.F. Trimm is professor, vice chair, and pediatric program director, Department of Pediatrics, University of South Alabama, Mobile, Alabama.

S. Guralnick is associate professor, director of graduate medical education, and designated institutional officer, Office of Academic Affairs, Winthrop University Hospital, Mineola, New York.

J.D. Mahan is professor, vice chair, and pediatric and pediatric nephrology fellowship program director, Department of Pediatrics, Nationwide Children's Hospital/Ohio State University, Columbus, Ohio.

K.A. Gifford is assistant professor and pediatric program director, Department of Pediatrics, Geisel School of Medicine at Dartmouth, Hanover, New Hampshire.

References

1 Institute of Medicine, Committee on the Governance Financing of Graduate Medical Education; Eden J, Berwick DM, Wilensky GR. Graduate Medical Education That Meets the Nation's Health Needs. Washington, DC: National Academies Press; 2014.
2 Balogh E, Miller BT, Ball J; Institute of Medicine, Committee on Diagnostic Error in Health Care. Improving Diagnosis in Health Care. Washington, DC: National Academies Press; 2015.
3 Lipner RS, Hess BJ, Phillips RL Jr. Specialty board certification in the United States: Issues and evidence. J Contin Educ Health Prof. 2013;33(suppl 1):S20–S35.
4 American Board of Medical Specialties. Steps toward initial certification and MOC. http://www.abms.org/board-certification/steps-toward-initial-certification-and-moc/. Accessed April 3, 2017.
5 Hauer KE, Ciccone A, Henzel TR, et al. Remediation of the deficiencies of physicians across the continuum from medical school to practice: A thematic review of the literature. Acad Med. 2009;84:1822–1832.
6 Nasca TJ, Philibert I, Brigham T, Flynn TC. The next GME accreditation system—Rationale and benefits. N Engl J Med. 2012;366:1051–1056.
7 Accreditation Council for Graduate Medical Education. Milestones. http://www.acgme.org/acgmeweb/tabid/430/ProgramandInstitutionalAccreditation/NextAccreditationSystem/Milestones.aspx. Published 2016. Accessed April 3, 2017.
8 Carraccio C, Gusic M, Hicks P, eds. The Pediatric Milestones Project—Competencies. Acad Pediatr. 2014;14(25 suppl):S13–S97.
9 Hicks PJ, Schumacher DJ, Benson BJ, et al. The pediatrics milestones: Conceptual framework, guiding principles, and approach to development. J Grad Med Educ. 2010;2:410–418.
10 Englander R, Hicks P, Benson B; Pediatric Milestone Project Working Group. Pediatrics milestones: A developmental approach to the competencies. J Pediatr. 2010;157:521–522.e1.
11 Englander R, Burke AE, Guralnick S, et al. The pediatrics milestones: A continuous quality improvement project is launched—Now the hard work begins! Acad Pediatr. 2012;12:471–474.
12 Schumacher DJ, Englander R, Hicks PJ, Carraccio C, Guralnick S. Domain of competence: Patient care. Acad Pediatr. 2014;14(2 suppl):S13–S35.
13 Guralnick S, Ludwig S, Englander R. Domain of competence: Systems-based practice. Acad Pediatr. 2014;14(2 suppl):S70–S79.
14 Schwartz A, Young R, Hicks PJ; Appd Learn. Medical education practice-based research networks: Facilitating collaborative research. Med Teach. 2014;38:1–11.
15 American Medical Association Fellowship and Residency Electronic Interactive Database. FREIDA Online. https://ama-assn.org/go/freida. Published 2016. Accessed April 3, 2017.
16 Accreditation Council for Graduate Medical Education. ACGME Data Resource Book: Academic Year 2013–2014 (released September 2014). http://www.acgme.org/About-Us/Publications-and-Resources/Graduate-Medical-Education-Data-Resource-Book/GraduateMedicalEducation/GraduateMedicalEducationDataResourceBook. Published 2014. Accessed April 3, 2017.
17 Agresti A. Categorical Data Analysis. 3rd ed. Hoboken, NJ: Wiley; 2013.
18 Conover WJ, Iman RL. Rank transformations as a bridge between parametric and nonparametric statistics. Am Stat. 1981;35:124–129.
19 Hosmer DW, Lemeshow S, Sturdivant RX. Applied Logistic Regression. 3rd ed. Hoboken, NJ: John Wiley & Sons; 2013.
20 Riebschleger MP, Haftel HM. Remediation in the context of the competencies: A survey of pediatrics residency program directors. J Grad Med Educ. 2013;5:60–63.
21 Guerrasio J, Garrity MJ, Aagaard EM. Learner deficits and academic outcomes of medical students, residents, fellows, and attending physicians referred to a remediation program, 2006–2012. Acad Med. 2014;89:352–358.
22 Yao DC, Wright SM. National survey of internal medicine residency program directors regarding problem residents. JAMA. 2000;284:1099–1104.
23 Silverberg M, Weizberg M, Murano T, Smith JL, Burkhardt JC, Santen SA. What is the prevalence and success of remediation of emergency medicine residents? West J Emerg Med. 2015;16:839–844.
24 Epstein RM, Siegel DJ, Silberman J. Self-monitoring in clinical practice: A challenge for medical educators. J Contin Educ Health Prof. 2008;28:5–13.
25 Blum RH, Boulet JR, Cooper JB, Muret-Wagstaff SL; Harvard Assessment of Anesthesia Resident Performance Research Group. Simulation-based assessment to identify critical gaps in safe anesthesia resident performance. Anesthesiology. 2014;120:129–141.
26 Accreditation Council for Graduate Medical Education; American Board of Pediatrics.

The Pediatrics Milestones Project. https://acgme.org/Portals/0/PDFs/Milestones/PediatricsMilestones.pdf. Published July 2015. Accessed April 10, 2017.
27 Choo KJ, Arora VM, Barach P, Johnson JK, Farnan JM. How do supervising physicians decide to entrust residents with unsupervised tasks? A qualitative analysis. J Hosp Med. 2014;9:169–175.
28 Sterkenburg A, Barach P, Kalkman C, Gielen M, ten Cate O. When do supervising physicians decide to entrust residents with unsupervised tasks? Acad Med. 2010;85:1408–1417.
29 Hauer KE, Oza SK, Kogan JR, et al. How clinical supervisors develop trust in their trainees: A qualitative study. Med Educ. 2015;49:783–795.
30 Sheu L, O'Sullivan PS, Aagaard EM, et al. How residents develop trust in interns: A multi-institutional mixed-methods study. Acad Med. 2016;91:1406–1415.
31 Choe JH, Knight CL, Stiling R, Corning K, Lock K, Steinberg KP. Shortening the miles to the milestones: Connecting EPA-based evaluations to ACGME milestone reports for internal medicine residency programs. Acad Med. 2016;91:943–950.
32 Bhatti NI, Ahmed A, Stewart MG, Miller RH, Choi SS. Remediation of problematic residents—A national survey. Laryngoscope. 2016;126:834–838.
33 Fargen KM, Drolet BC, Philibert I. Unprofessional behaviors among tomorrow's physicians: Review of the literature with a focus on risk factors, temporal trends, and future directions. Acad Med. 2016;91:858–864.
34 Papadakis MA, Teherani A, Banach MA, et al. Disciplinary action by medical boards and prior behavior in medical school. N Engl J Med. 2005;353:2673–2682.
35 Papadakis MA, Arnold GK, Blank LL, Holmboe ES, Lipner RS. Performance during internal medicine residency training and subsequent disciplinary action by state licensing boards. Ann Intern Med. 2008;148:869–876.

Appendix 1

Discrimination Between Marginal/Unsatisfactory and Satisfactory Pediatric Residents for Different Subcompetencies by Level of Training Corrected for Multiple Comparisons With Bonferroni Method, From a Multi-Institutional Study of Clinical Skills Progression by Pediatric Milestones, 2013–2014[a]

	PGY1		PGY2		PGY3	
Subcompetency	**AUC**	**99.92% CI**	**AUC**	**99.92% CI**	**AUC**	**99.92% CI**
PC1: Gather essential and accurate information about the patient	0.87	0.74–0.99	0.61	0.24–0.98	0.62	0.13–1.00
PC2: Organize and prioritize responsibilities to provide patient care that is safe, effective, and efficient	0.91	0.78–1.00	0.68	0.38–0.98	0.65	0.17–1.00
PC3: Provide transfer of care that ensures seamless transitions	0.90	0.79–1.00	0.74	0.48–1.00	0.64	0.15–1.00
PC4: Make informed diagnostic and therapeutic decisions that result in optimal clinical judgment[b]	0.90	0.79–1.00	0.70	0.37–1.00	0.76	0.46–1.00
PC5: Develop and carry out management plans	0.89	0.73–1.00	0.65	0.32–0.97	0.65	0.23–1.00
MK: Locate, appraise, and assimilate evidence from scientific studies related to their patients' health problems	0.86	0.68–1.00	0.60	0.28–0.93	0.47	0.00–1.00
PBLI1: Identify strengths, deficiencies, and limits in one's knowledge and expertise	0.89	0.75–1.00	0.64	0.27–1.00	0.62	0.17–1.00
PBLI2: Identify and perform learning activities to guide personal and professional development	0.86	0.70–1.00	0.65	0.28–1.00	0.60	0.10–1.00
PBLI3: Systematically analyze practice using quality improvement methods, and implement changes with the goal of practice improvement	0.80	0.59–1.00	0.74	0.36–1.00	0.82	0.50–1.00
PBLI4: Incorporate formative evaluation feedback into daily practice	0.85	0.65–1.00	0.62	0.25–0.99	0.73	0.28–1.00
ICS1: Communicate effectively with patients, families, and the public, as appropriate, across a broad range of socioeconomic and cultural backgrounds	0.83	0.67–1.00	0.51	0.12–0.90	0.63	0.13–1.00
ICS2: Demonstrate the insight into emotion and human response to emotion that allows one to appropriately develop and manage human interactions	0.82	0.62–1.00	0.66	0.37–0.96	0.77	0.41–1.00
Prof1: Humanism, compassion, integrity, and respect for others; based on the characteristics of an empathetic practitioner	0.79	0.58–0.99	0.55	0.25–0.85	0.62	0.19–1.00
Prof2: Professionalization: A sense of duty and accountability to patients, society, and the profession	0.79	0.56–1.00	0.61	0.27–0.96	0.70	0.28–1.00
Prof3: Professional conduct: High standards of ethical behavior which includes maintaining appropriate professional boundaries	0.76	0.51–1.00	0.58	0.28–0.89	0.79	0.21–1.00
Prof4: Self-awareness of one's own knowledge, skill, and emotional limitations that lead to appropriate help-seeking behaviors	0.87	0.67–1.00	0.63	0.24–1.00	0.58	0.00–1.00
Prof5: Trustworthiness that makes colleagues feel secure when one is responsible for the care of patients	0.84	0.65–1.00	0.73	0.42–1.00	0.66	0.20–1.00
Prof6: The capacity to accept that ambiguity is part of clinical medicine and to recognize the need for and to utilize appropriate resources in dealing with uncertainty	0.86	0.67–1.00	0.63	0.27–0.99	0.62	0.14–1.00
SBP1: Coordinate patient care within the health care system relevant to their clinical specialty[b]	0.87	0.71–1.00	0.63	0.35–0.91	0.67	0.33–1.00
SBP2: Advocate for quality patient care and optimal patient care systems	0.77	0.50–1.00	0.63	0.19–1.00	0.87	0.56–1.00
SBP3: Work in interprofessional teams to enhance patient safety and improve patient care quality[b]	0.83	0.62–1.00	0.62	0.23–1.00	0.61	0.15–1.00

Abbreviations: PGY indicates postgraduate year; AUC, area under receiver operating characteristic curve; PC, patient care; MK, medical knowledge; PBLI, practice-based learning and improvement; ICS, interpersonal and communication skills; Prof, professionalism; SBP, systems-based practice.

[a]The authors calculated asymptotic 99.92% CIs for the AUC (area under the receiver operating characteristic curve) based on Somers D. A Bonferroni-corrected significance threshold of .0007963 was used to account for the 63 comparisons. C-statistics can be interpreted as the probability that a randomly selected trainee identified as marginal/unsatisfactory has a lower milestone score than a satisfactory trainee. A C-statistic of 0.5 indicates that a subcompetency is no better than chance at discriminating between M/U and S residents, whereas a C-statistic of 1 indicates that a subcompetency perfectly discriminates between M/U and S residents. C-statistic values of 0.7–0.8 indicate acceptable discrimination, values of 0.8–0.9 indicate excellent discrimination, and values ≥ 0.9 indicate outstanding discrimination.

[b]These subcompetencies have only four milestone levels.