

The Radar Graph: the Development of an Educational Tool to Demonstrate Resident Competency

DREW M. KEISTER, MD, FAAFP
DANIEL LARSON, MS
JULIE DOSTAL, MD
JAY BAGLIA, PhD

Abstract

Background Despite the movement toward competency-based assessment by accrediting bodies in recent years, there is no consensus on how to best assess medical competence. Direct observation is a useful tool. At the same time, a comprehensive assessment system based on direct observation has been difficult to develop.

Intervention We developed a system that translates data obtained from checklists of observed behaviors completed during educational activities, including direct observation of clinical care, into a graphic tool (the “radar graph”) usable for both formative and summative assessment. Using unique, observable behaviors to evaluate levels of competency on the Dreyfus scale, we assessed resident performance in 6 learning sites within our residency. Data are

represented on a radar graph, which residents and faculty used to recognize both strengths and areas for growth to guide educational planning for the individual learner.

Results Initial data show that the radar graphs have construct validity because the development process accurately reflects the desired construct, assessors were adequately trained, and the radar graphs demonstrated resident growth over time. A form completion rate of 90% for >1500 disseminated assessments suggests the feasibility of our process.

Conclusions The radar graph is a promising tool for use in resident feedback and competency assessment. Further research is needed to determine the full utility of the radar graphs, including a better understanding of the tool’s reliability and construct validity.

Introduction

Since the Accreditation Council for Graduate Medical Education (ACGME) shifted from process to outcomes or competency-based accreditation in 1999,¹ there has been discussion about the preferred assessment methods but no consensus has been reached.² While some studies have focused on the assessment of residency programs,³ more scholarly attention has centered on resident assessment.^{2,4–7} Focus on the performance of individual residents

demonstrates a need for competency-based education and assessment.

One useful assessment tool is the behavioral checklist, which offers good interrater reliability and immediate formative feedback to learners.^{7,8} When checklists are used in direct observation of clinical care, a wealth of assessment data can be obtained. However, a comprehensive system based on direct observation is difficult to achieve^{2,4,8} and requires frequent assessment in multiple contexts.^{2,8,9} Behavioral checklists require that competencies be distilled into observable behaviors.⁸ At the same time, long lists of behaviors created by a deconstruction of the competencies are difficult to compile into a framework that supports learner growth.¹⁰

We concluded that learners would benefit from a system that translates data obtained from direct observation checklists into a graphic form, usable for both formative and summative assessment. Such a system would avoid the flaws of checklists that deconstruct competencies into pieces by compiling multiple behaviors for each competency into a single visual tool or “radar graph.” The tool would help learners better understand their formative and summative assessments and become a focus for a reflective practice¹¹ that could contribute to learner-developed educational goals and drive a desire for self-improvement.

Drew M. Keister, MD, FAAFP, is Clerkship Director at Lehigh Valley Health Network Family Medicine Residency (LVHN FMR); **Daniel Larson, MS**, is a Research Assistant at LVHN FMR; **Julie Dostal, MD**, is Vice Chair of Education and Residency Director at LVHN FMR; and **Jay Baglia, PhD**, is Assistant Professor of Communications Studies at Kutztown University.

Funding: Funding for this work was provided by the Dorothy Rider Pool Health Care Trust.

The authors would like to acknowledge Elissa Foster, PhD, and Susan Mathieu, MD, for their contributions to the system development and review of this manuscript, and William Miller, MD, for his visionary leadership.

Corresponding author: Drew Keister, MD, Department of Family Medicine, 17th and Chew Streets, School of Nursing, 1st Floor, Allentown, PA 18105, drew_m.keister@lvhn.org

Received: July 12, 2011; revision received: November 8, 2011; accepted: November 15, 2011.

DOI: <http://dx.doi.org/10.4300/JGME-D-11-00163.1>

We describe the methods for system creation, share a sample radar graph, and discuss a preliminary assessment of the graphs' construct validity using data about the content, response process, and internal structure of our tool. We created the radar graphs for combined formative/summative assessment based on direct observation of resident behavior. The graphs facilitate the assessment of residents' ability to function as complete physicians, integrating specialty-specific competencies in actual clinical care. We also sought to validate the radar graph tool, beginning with a demonstration of construct validity of the assessed construct.¹²

Methods

Setting

The Lehigh Valley Health Network Family Medicine Residency trains 6 residents in each class year. In 2007, we began a transformative journey as 1 of 14 US family medicine residencies in the Preparing the Personal Physician for Practice (P⁴) project. A key portion of our P⁴ innovation was to develop and implement a competency-based assessment system.

Radar Graph Creation

We previously reported the process of generating our developmentally appropriate assessment system.¹³ Briefly, we first customized the ACGME core competencies to fit our learner population and philosophies specific to family medicine. Our program-specific family medicine (FM) competencies are listed in the BOX. We used a narrative process to define our competencies, thereby generating standard statements for each. Next, we tiered the standard statements according to the Dreyfus model for skill acquisition,^{3,14} assigning standards to novice, advanced beginner, competent, and proficient levels. Residents and faculty generated observable behaviors within the standards.¹³

We used these observable behaviors to create assessment forms in 6 settings where residents demonstrated competency relevant to independent practice in FM.¹³ FIGURE 1 describes the process for form completion and data conversion into radar graphs. Data on resident assessments collected via New Innovations (Uniontown, OH) was downloaded into a database that displays the data at multiple levels (individual resident, postgraduate year [PGY] class, entire program, a single competency, and other levels). Using a conceptual framework described by Nasca,¹⁵ we created "radar graphs" of the percentage of observations during which residents meet each standard. Residents and their assessment teams use the graphs to generate individualized education plans. The

What was known

Behavioral checklists are a useful assessment tool that offer useful formative feedback to learners. At the same time, development of a comprehensive evaluation system based on direct observation is more challenging.

What is new

Development of a "radar graph" showing observable behaviors related to the ACGME competencies facilitated assessment of family medicine resident performance across 6 learning sites. The graphs showed strengths and areas for growth to guide educational planning for learners.

Limitations

Single-site, single-specialty study, and small sample may limit generalizability. Acknowledged limitations of the assessment tool and conversion of nominal to ordinal data may hamper analysis and interpretation of the data.

Bottom line

The "radar graph" is a promising tool that can be populated with existing data for use in competency assessment and resident feedback.

implementation of these forms began in June 2009, and the final data set was added in May 2011.

Methods for Validity Analysis

Content evidence examines how well the content of the assessment matches the desired *construct*.¹² Above, we described the process for developing our assessment tools. This description is both the method for tool creation and content evidence demonstrating the relationship between our construct and the assessment content. We gathered evidence for construct validity from 5 sources: content, response process, internal structure, relation to other variables, and consequences.¹²

Response process evidence describes the link between the construct and the *thought processes* of assessors.¹² We report the process used for assessor training and for data reporting. Additionally, we examined the completion rate for all assigned assessments.

Internal structure evidence speaks to the reliability of the assessment tool.¹² One measure of internal structure is provided by a pattern of responses predicted by the construct.¹² Our construct predicts that residents' competency

BOX LEHIGH VALLEY HEALTH NETWORK'S FAMILY MEDICINE COMPETENCIES

Comprehensive Care (CC)
Community Health Partnership (CHP)
Information Literacy and Knowledge Creation (IL/KC)
Leadership and Change Management (LCM)
Lifelong Learning (LL)
Relationship-Centered Care (RCC)
Self Care (SC)

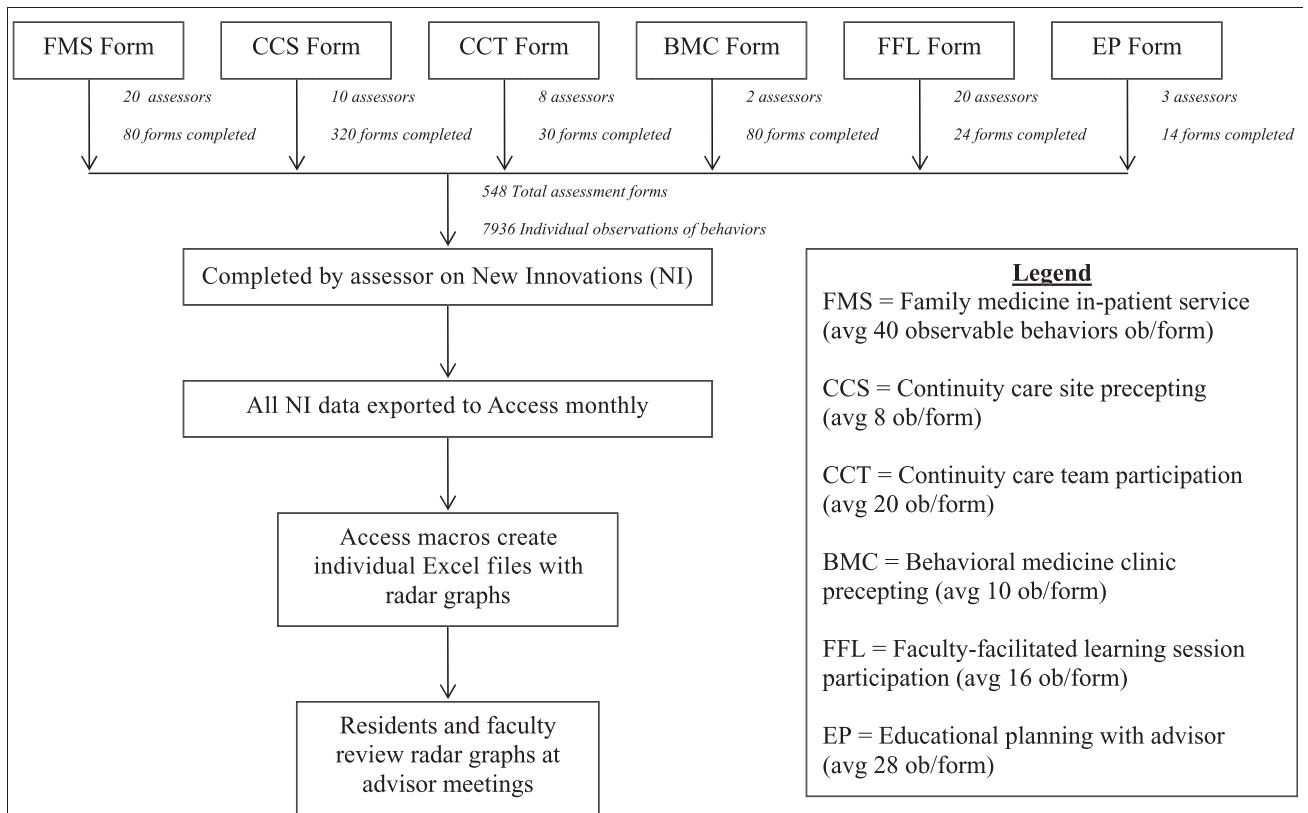


FIGURE 1 | **FLOW SHEET OF ASSESSMENT FORM COMPLETION**

- The assessment forms and a grid delineating all competencies, standard statements and observable behaviors are provided as online content.
- A full description of when forms are completed is available in Baglia et al.¹³
- Number of assessors and assessments completed is an estimate of completion over the course of a 3-year family medicine residency.

grows over time. To demonstrate this prediction, we visually compared the combined graphs of resident cohorts (by PGY) to look for visually perceptible growth. To show this change at the individual level, we developed the “composite competency rating” (CCR). This single number between 0 and 10 examines a resident’s overall competency by averaging numerical values given to the responses. A value of 0 would represent a radar graph that looked like a small dot, while 10 would represent a perfect circle filling the entire graph.

Results

The Radar Graph

The primary result of the assessment component of our P⁴ intervention is the radar graph. FIGURE 2 represents the “target” competency of a graduating resident. Although this target is not realistic for actual residents, it is useful to conceptualize the meaning of the figure. The radar graph, a visual representation of competency, is broken up into 7 sections, one for each FM competency.¹³ Each section consists of the standards that make up the competency. These standards are arranged clockwise in their

competency by Dreyfus level (novice to proficient). The size of each section is based upon the number of standards that make up that competency. Each data point is a composite of the observable behaviors derived from that standard.

Following one standard from the center of the graph outward, each point represents the frequency with which the resident is observed performing that behavior. As the resident performs the behaviors within a standard more frequently, the data point moves outward, thus creating a graph with a larger shaded area. A fully proficient physician’s graph would be a perfect circle. Note that the “target” graduate of our program would be competent (but not always proficient) in the FM competencies.

FIGURE 3, another sample graph, shows the target for progression from PGY-1 to PGY-3. This comparison of 3 target cohorts shows predicted resident growth over time.

Assessors were trained in several ways. Residents and faculty were integral in writing the observable behaviors. Thus, the creation process was also an assessor development technique. Additionally, assessors received guidance via e-mail and participatory didactic sessions on each form. Monthly faculty development workshops include training

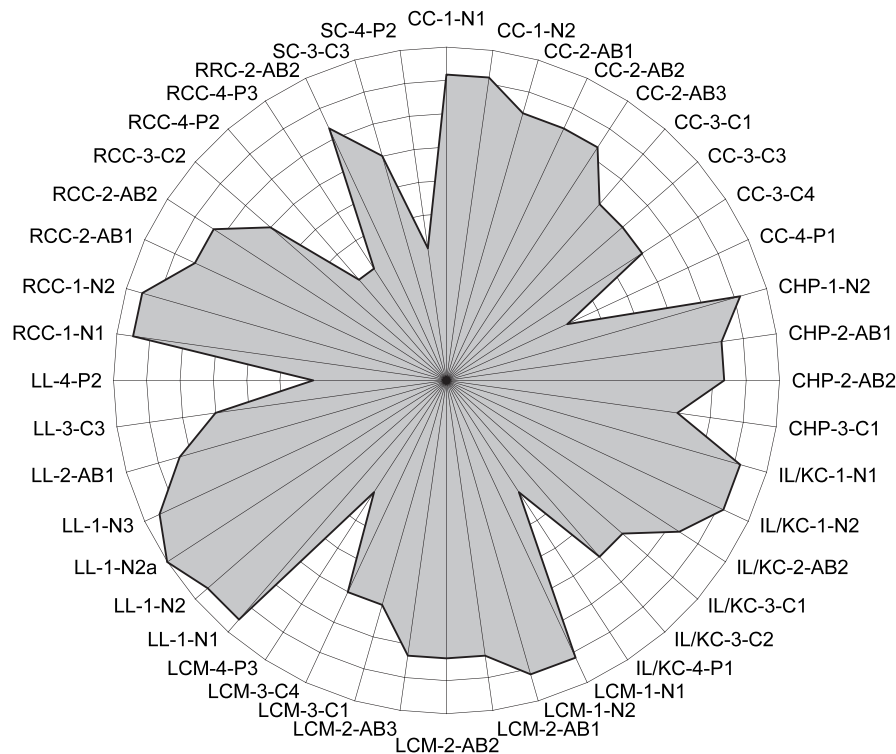


FIGURE 2 THEORETICAL TARGET RESIDENT GRAPH

Each axis of this circular radar graph represents one standard statement from our competency table (full table available as supplementary online material). The axes are labeled as "Competency abbreviation-Dreyfus level number-Dreyfus abbreviation and standard number" (eg, "CC-1-N1" is comprehensive care competency, novice Dreyfus level, first standard statement; and "LCM-2-AB2" is leadership change management competency, advanced beginner Dreyfus level, second standard statement).

about the assessment process, as does our annual Residents as Teachers and Leaders curriculum.

Data from completed assessment forms is reported by New Innovations directly to the assessed resident, the residents' advisor, and the program director. The program manager is the only other person who has data access. FIGURE 1 describes the process for data transfer from New Innovations to the radar graphs.

A total of 1372 of 1508 assessment forms (90%) have been completed in New Innovations. No incentive was given to encourage completion.

FIGURE 4 shows the growth comparison of 3 resident cohorts. The larger graph for each PGY cohort shows the predicted growth. Additionally, the TABLE shows residents' CCRs, which increase over time. Thus, the construct-predicted growth over time is seen both at the cohort and individual levels.

Discussion

The radar graph offers a new tool for resident feedback and competency assessment. The graphs have content evidence of validity through use of the ACGME competencies as their foundation and sound educational theory in their

development.^{7,8,13,16} Assessors received appropriate training, data were reported confidentially and reliably, and assessment form completion was excellent. Examining the progression of residents in the TABLE and in FIGURE 4, we confirmed the prediction that competence progresses with time. This suggests that our assessment process quantifies the time-based competency assumption that predated the ACGME Outcomes Project.²

Beyond validity, the radar graph has potential to make feedback more accessible to learners. Residents arrive to training programs with unique skill sets and acquire new skills at different rates.^{17,18} The radar graph identifies when residents develop competency on differing trajectories (eg, the learner who demonstrates competent-level medical knowledge at the beginning of residency, but novice-level interpersonal communication). Thus, learners and their mentors can partner to improve weaknesses and bolster strengths.

Additionally, using direct observation in multiple settings avoids the decontextualization of a competency,⁹ a frequent problem in residency assessment. Manipulations of the graph can create representations of residents' performance at each assessment site. Therefore, residents can develop reflective practices around self-assessment of

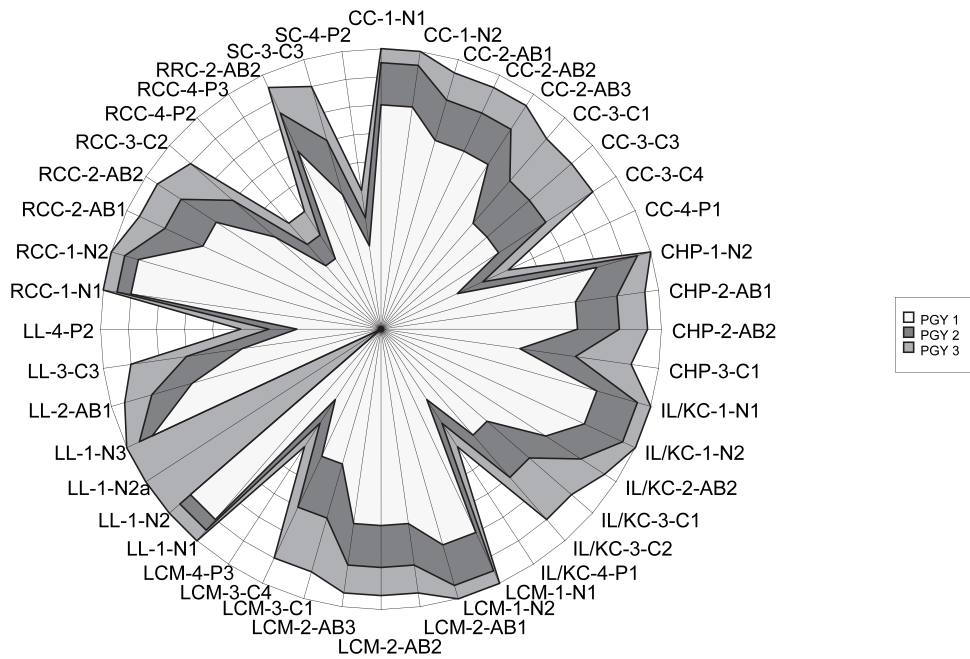


FIGURE 3 THEORETICAL TARGET PROGRESSION ACROSS RESIDENCY COHORTS

competency in differing settings. Thus, our learners gain help in recognizing when they are in an environment where they are less competent.¹⁰

Our competencies are not mutually exclusive from those of the ACGME. All of our competencies can be understood within the ACGME model.¹³ We can create radar graphs using ACGME competencies. However, in generating FM competencies we facilitated our community's faculty/resident development, creating common understanding and shared meaning of the competencies. The process of creating the system produced engagement and skill-building in the components of the system. This formed the backbone of our evaluator development. The "buy-in" created therein contributes to the high form completion rate.

Our approach has several limitations. The first limitation is the conversion of nominal to ordinal data to permit radar graph creation and statistical analysis. As a result of this, statistical differences may not reflect actual differences within the data. Also, we cannot use Cronbach's alpha to measure reliability. However, Krippendorff's alpha may provide meaningful analysis of reliability across multiple evaluators.¹⁹ In the future, we plan to use this measure with multiple evaluators assessing identical resident performances to further demonstrate the reliability of our system.

The assessment forms also have limitations. Informal feedback demonstrates a need for improved language describing our specific behaviors. We have reconstituted

several forms, making them more uniform to evade evaluator confusion.

Conclusion

We created functioning radar graphs based on composite data of observable behaviors from 6 learning sites that allow residents and their advisors to visually identify areas of strength and growth opportunities. The radar graph is a generalizable tool for resident assessment. Our process could be recreated in residency programs of any specialty and could examine constructs other than ours. Established tools that use direct observation could be used to create radar graphs. Accrediting bodies and/or residency directors from each specialty could identify sites within their residencies where multiple competencies could be witnessed in toto and develop observable behaviors within these sites. Programs could use the ACGME competencies directly or to develop linked specialty-specific competencies.

Further research is needed to expand our understanding of the construct validity of this tool.¹² External validity will be tested by correlating the results of radar graphs with standard intermediate markers for competence, such as the patient-empowerment index, patient enablement instrument, in-training scores, happiness and life satisfaction, Jefferson scale of lifelong learning, and stewardship. "Expert" and perhaps "master" Dreyfus-level behaviors will be identified.^{20,21} Finally, the ease of use and accessibility of this method need to be studied.

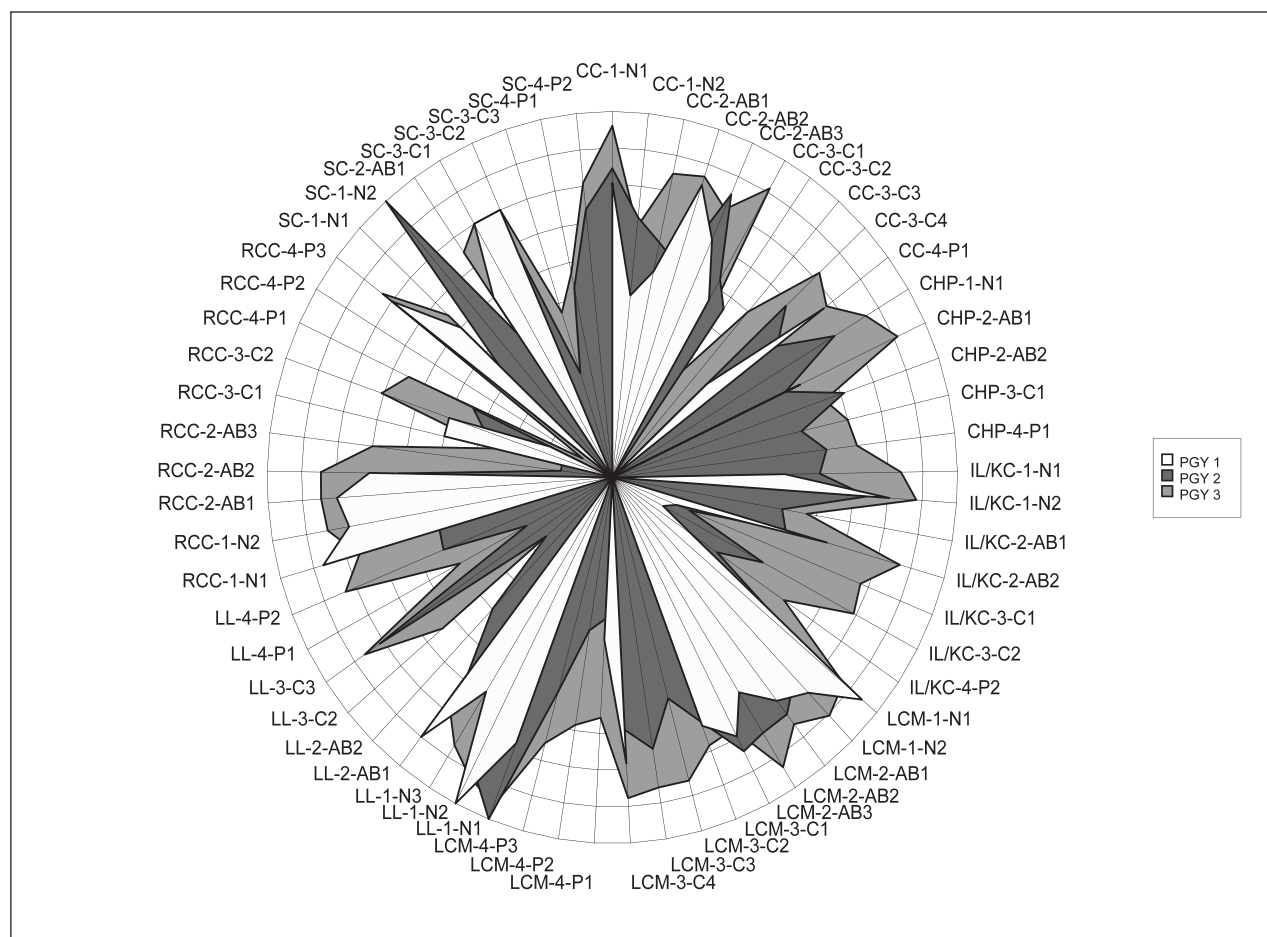


FIGURE 4 ACTUAL PROGRESSION ACROSS RESIDENCY COHORTS

TABLE	COMPREHENSIVE COMPETENCY RATING (CCR) FOR EACH RESIDENT BY SEMESTER AND PGY		
CCR	First 6-month period	Second 6-month period	Final 6-month period
PGY-3 Resident A	7.19	7.660	8.36
PGY-3 Resident B	3.71	6.71	8.00
PGY-3 Resident C	6.29	7.95	8.26
PGY-3 Resident D	4.69	6.09	7.85
PGY-2 Resident E	6.04	5.98	7.59
PGY-2 Resident F	6.00	5.65	9.25
PGY-2 Resident G	3.53	6.91	7.54
PGY-2 Resident H	5.55	6.28	8.21
PGY-2 Resident I	3.99	6.35	7.35
PGY-2 Resident J	3.59	6.44	7.88
PGY-2 Resident K	6.60	7.18	8.51

CCR is the "composite competency rating" (numbered between 0 and 10). CCR is calculated by averaging numerical values assigned to the responses (ie, observed = 10, not observed = 0, no chance to observe = response not counted).

This table represents all of the data on residents from these two classes. Although we generally have 6 residents per class year, 2 residents in this PGY-3 cohort left the residency because of academic difficulties unrelated to this assessment system. The PGY-2 cohort had 1 additional resident who filled the spot of one of the PGY-3 residents who left the program. The PGY-1 cohort was not present for the first year of assessment, so their CCRs did not contribute to this analysis.

References

- 1 Accreditation Council for Graduate Medical Education. ACGME Common Program Requirements. http://www.acgme.org/acWebsite/dutyHours/dh_dutyhoursCommonPRO7012007.pdf. Accessed March 5, 2012.
- 2 Lurie SJ, Mooney CJ, Lyness JM. Measurement of the General Competencies of the Accreditation Council for Graduate Medical Education: a systematic review. *Acad Med*. 2009;84(3): 301–309.
- 3 Batalden P, Leach D, Swing S, Dreyfus H, Dreyfus S. General competencies and accreditation in graduate medical education. *Health Affairs*. 2002; 21(5):103–111.
- 4 Brasel KJ, Bragg D, Simpson DE, Weigelt JA. Meeting the Accreditation Council for Graduate Medical Education competencies using established residency training program assessment tools. *American J Surg*. 2004;188:9–12.
- 5 Epstein RM, Hundert EM. Defining and assessing professional competence. *JAMA*. 2002; 287(2):226–235.
- 6 Shayne P, Gallahue F, Rinnert S, Anderson CL, Hern G, Katz E. Reliability of a core competency checklist assessment in the emergency department: the standardized direct observation assessment tool. *Acad Emerg Med*. 2006;13(7):727–732.
- 7 Accreditation Council for Graduate Medical Education and American Board of Medical Specialties. Toolbox of assessment methods: version 1.1, September 2000. http://www.ihl.org/offerings/IHIOpenSchool/resources/Assets/Tools%20-%20ToolboxofAssessmentMethodstoEvaluateResidents_28101d1-83fo-4765-88f8-bogbad31f852/Toolbox.pdf. Accessed March 22, 2012.
- 8 Swing SR. Assessing the ACGME General Competencies: general considerations and assessment methods. *Acad Emerg Med*. 2002;9(11):1278–1288.
- 9 Lingard L. What we see and don't see when we look at "competence": notes on a god term. *Adv Health Sci Educ Theory Pract*. 2009;14: 625–628.
- 10 Frank JR, Mungroo R, Ahmad Y, Wang M, DeRossi S, Horsley T. Toward a definition of competency-based education in medicine: a systematic review of published definitions. *Med Teach*. 2010;32:631–637.
- 11 Mann K, Gordon J, MacLeod A. Reflection and reflective practice in health professions education: a systematic review. *Adv Health Sci Educ Theory Pract*. 2009;14:595–621.
- 12 Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. *Am J Med*. 2006;119:166e7–166e16.
- 13 Baglia J, Foster E, Dostal J, Keister K, Biery N, Larson D. Generating developmentally appropriate competency assessment at a family medicine residency. *Fam Med*. 2011;43(2):90–98.
- 14 Dreyfus SE, Dreyfus HL. *Mind Over Machine: The Power Of Human Intuition And Expertise In The Era Of The Computer*. New York: The Free Press; 1986.
- 15 Nasca TF. Where will the "milestones" take us? The next accreditation system. *ACGME Bulletin*. 2008; September.
- 16 Williams RG, Klamen DA, McGaghie WC. Cognitive, social and environmental sources of bias in clinical performance ratings. *Med Teach*. 2003;15(4):270–292.
- 17 ten Cate O, Snell L, Mann K, Vermunt J. Orienting teaching toward the learning process. *Acad Med*. 2004;79(3):219–228.
- 18 Carraccio, CL, Benson BJ, Nixon LJ, Derstine PL. From the educational bench to the clinical bedside: translating the Dreyfus developmental model to the learning of clinical skills. *Acad Med*. 2008;83(8):761–767.
- 19 Hayes AF, Krippendorff K. Answering the call for a standard reliability measure for coding data. *Commun Methods Meas*. 2007;1(1):77–89.
- 20 Peña A. The Dreyfus model of clinical problem-solving skills acquisition: a critical perspective. *Med Educ Online*. 2010 Jun 14;15. doi: 10.3402/meo.v15i0-4846.
- 21 Dreyfus HL. *On the Internet: Thinking in Action*. London: Routledge; 2001.