

Test of Integrated Professional Skills: Objective Structured Clinical Examination/Simulation Hybrid Assessment of Obstetrics-Gynecology Residents' Skill Integration

ABIGAIL FORD WINKEL, MD
COLLEEN GILLESPIE, PHD
MARISSA T. HIRUMA, MA
ALICE R. GOEPFERT, MD
SONDRA ZABAR, MD
DEMIAN SZYLD, MD, EdM

Abstract

Background Assessment of obstetrics-gynecology residents' ability to integrate clinical judgment, interpersonal skills, and technical ability in a uniform fashion is required to document achievement of benchmarks of competency. An observed structured clinical examination that incorporates simulation and bench models uses direct observation of performance to generate formative feedback and standardized evaluation.

Methods The Test of Integrated Professional Skills (TIPS) is a 5-station performance-based assessment that uses standardized patients and complex scenarios involving ultrasonography, procedural skills, and evidence-based medicine. Standardized patients and faculty rated residents by using behaviorally anchored checklists. Mean scores reflecting performance in TIPS were compared across competency domains and by developmental level (using analysis of variance) and then compared to standard faculty clinical evaluations (using Spearman

ρ). Participating faculty and residents were also asked to evaluate the usefulness of the TIPS.

Results Twenty-four residents participated in the TIPS. Checklist items used to assess competency were sufficiently reliable, with Cronbach α estimates from 0.69 to 0.82. Performance improved with level of training, with wide variation in performance. Standard faculty evaluations did not correlate with TIPS performance. Several residents who were rated as average or above average by faculty performed poorly on the TIPS (> 1 SD below the mean). Both faculty and residents found the TIPS format useful, providing meaningful evaluation and opportunity for feedback.

Conclusions A simulation-based observed structured clinical examination facilitates observation of a range of skills, including competencies that are difficult to observe and measure in a standardized way. Debriefing with faculty provides an important interface for identification of performance gaps and individualization of learning plans.

Abigail Ford Winkel, MD, is Assistant Professor and Co-Associate Residency Program Director, Department of Obstetrics and Gynecology, New York University School of Medicine; **Colleen Gillespie, PhD**, is Assistant Professor, Division of General Internal Medicine, New York University School of Medicine; **Marissa T. Hiruma, MA**, is Program Manager, Graduate Medical Education, New York University School of Medicine; **Alice R. Goepfert, MD**, is Director, OBGYN Residency Program and Director of Education, Department of Obstetrics and Gynecology, University of Alabama at Birmingham; **Sondra Zabar, MD**, is Associate Professor, Division of General Internal Medicine, New York University School of Medicine; and **Demian Szyld, MD, EdM**, is Assistant Professor, Department of Emergency Medicine, New York University School of Medicine, and Associate Medical Director, New York Simulation Center of the Health Sciences.

Funding: The authors report no external funding source for this study.

Findings were presented at the CREOG and APGO Annual Meeting, February 27–March 2, 2013, Phoenix, Arizona.

The authors would like to thank the following individuals at the New York University School of Medicine for their contribution to the development and implementation of the Test of Integrated Professional Skills: Dr Veronica Lerner, Dr Sigrid Tristan, Ms Ginny Drda, and Ms Dorice Vieira.

Corresponding author: Abigail Ford Winkel, MD, 550 First Avenue, NBV 9E5, New York, NY 10016, 212.263.8683, abigail.winkel@nyumc.org

Introduction

Physician competence is difficult to capture,¹ and deconstructing competence into measurable segments may result in loss of the overall sense of a physician's ability.^{2–6} The heterogeneous nature of residents' on-the-job training⁷ necessitates standardized assessment tools that mitigate inherent rater bias.^{8,9} Reliable evaluation tools are essential for Milestone assessment in the Next Accreditation System for graduate medical education.⁹

Objective structured clinical examination (OSCE) tools generate information about several dimensions of performance, including those difficult to evaluate by traditional means. OSCEs have high validity and educational impact in

Received February 15, 2013; revision received May 13, 2013; accepted July 29, 2013.

DOI: <http://dx.doi.org/10.4300/JGME-D-13-00055.1>

graduate medical education.^{10–19} Trained standardized patient raters provide important measures of performance.^{4,13,20} Educators in procedural-based specialties have developed a proliferation of bench models and simulated environments to evaluate procedural skills.^{3,21} Skill assessments for procedures using global rating scales and checklists display moderate interrater reliability for performance.^{22,23}

Obstetrician-gynecologists must be able to integrate a wide range of skills, but there are limited tools for assessing integrated skills. The Test of Integrated Professional Skills (TIPS) requires residents to interact with standardized patients in clinical scenarios and perform skills on models for forceps delivery, transvaginal ultrasonography, intra-uterine device placement, and laparoscopic suturing. The aim of TIPS is to assess integration of skills on an individual resident. Faculty observers can identify gaps in resident ability and provide immediate feedback and directed teaching. We describe the implementation of the TIPS examination and a comparison of resident performance data from the examination with faculty data obtained during regular rotation evaluations.

Methods

The TIPS examination was developed to assess performance across Accreditation Council for Graduate Medical

What was known

Performance in practice requires physicians to integrate clinical judgment, technical ability, and interpersonal skills.

What is new

A 5-station assessment using hybrid simulation that combined standardized patients and simulated scenarios to test a higher order and integrated skills.

Limitations

Single-institution study limits generalizability. Some skills were assessed only once; the 5 scenarios do not provide a comprehensive assessment of a resident's skills.

Bottom line

A simulation-based, objective structured clinical examination facilitated assessment of competencies that are difficult to observe and measure in a standardized way. Debriefing allowed for identification of performance gaps.

Education (ACGME) competencies in an integrated manner for obstetrics-gynecology residents. We developed 5 OSCE stations (TABLE 1) that combined clinical interaction using a standardized patient with a specific skill demonstration based on a format piloted with interns in 2011.²⁴ Content validity for the skill stations and checklists was developed through consultation with 2 obstetrics-gynecology faculty members and 1 simulation expert from

TABLE 1	ASSESSMENT MATRIX FOR THE TEST OF INTEGRATED PROFESSIONAL SKILLS				
Competency	Elements	No. of Cases Assessed	No. of Checklist Items	Internal Consistency (Cronbach α)	Rater
Practice-based learning and improvement	Evidence-based medicine	1	5	0.74	Faculty
Patient care: procedural skills	Ultrasonography	4	38	0.67	
	Laparoscopic suturing				
	Intrauterine device placement				
	Forceps delivery				
Patient care: assessment, counseling, treatment plan	History gathering	3	12	0.71	Standardized patient
	Counseling				
	Treatment plan/next steps				
Communication skills	Information gathering	5	60	0.82	
	Relationship development				
	Patient education and counseling				
Professionalism	Informed consent	4	24	0.73	
	Difficult situation				
	Breaking bad news				
Systems-based practice	Interprofessional collaboration	1	3	0.69	

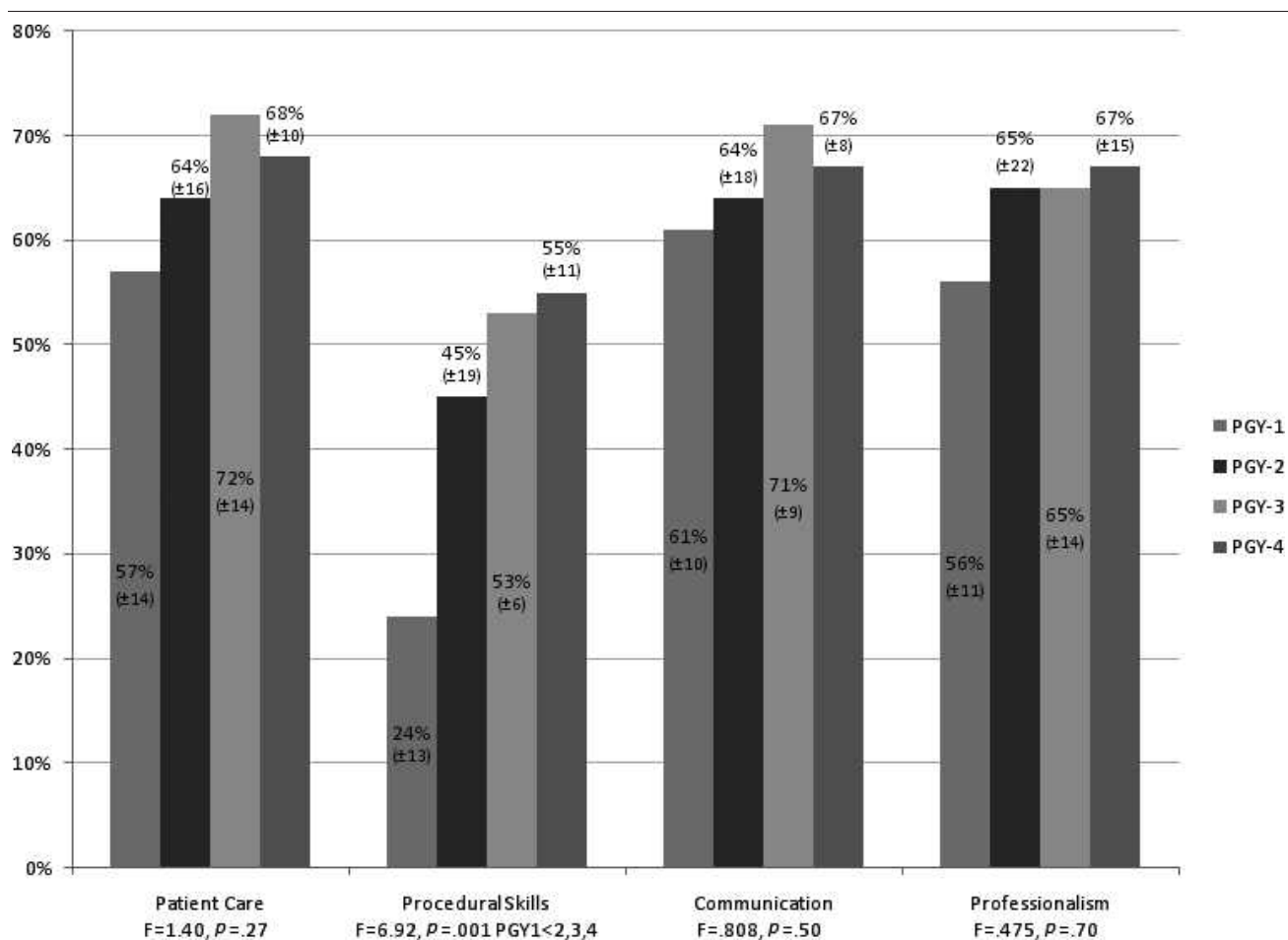


FIGURE 1 | RESIDENT PERFORMANCE ON TEST OF INTEGRATED PROFESSIONAL SKILLS

Mean scores for residents by postgraduate year (PGY) based on percentage of test items “well-done” by competency with standard deviation. Scores improve with PGY year, but show large variations in performance. Significance of differences between means by PGY was assessed through 1-way analysis of variance with Bonferroni post hoc pairwise comparisons.

the New York Simulation Center for the Health Sciences. The medical literature station was developed in consultation with a medical librarian.

Standardized patients were trained as raters using a behaviorally anchored checklist. Faculty members scoring stations were experts in the skill area being addressed. Checklists incorporated items from OSCEs used over a decade of experience at our institution in medical education at all levels.^{13,25–28} Standardized patients rated interpersonal and cognitive skills; faculty rated technical skills.

Each station was 16 minutes long, followed by 5 minutes of feedback with faculty observers. Residents and faculty were surveyed with a paper questionnaire. Residents received summative evaluation in the form of a TIPS Report Card during semiannual evaluations with the program director. Items for each checklist were categorized by ACGME competencies across stations. Cronbach α was

calculated for all items to determine their reliability (internal consistency; TABLE 1). Results from the TIPS examination were compared to online rotation evaluations completed by faculty during the 2011–2012 academic year. Each resident had between 15 and 32 ratings per year that were averaged to arrive at a mean rating. These were averaged across the year to account for variation in timing of rotation blocks. We assessed the association between standardized z scores from TIPS performance and faculty rotation ratings within each competency by using Spearman correlation coefficient. The sample size was small, limiting its power to detect a small difference between the 2 assessments.

The Institutional Research Board at New York University School of Medicine approved this study, which used the Research on Medical Education Outcomes Registry for deidentified educational data.

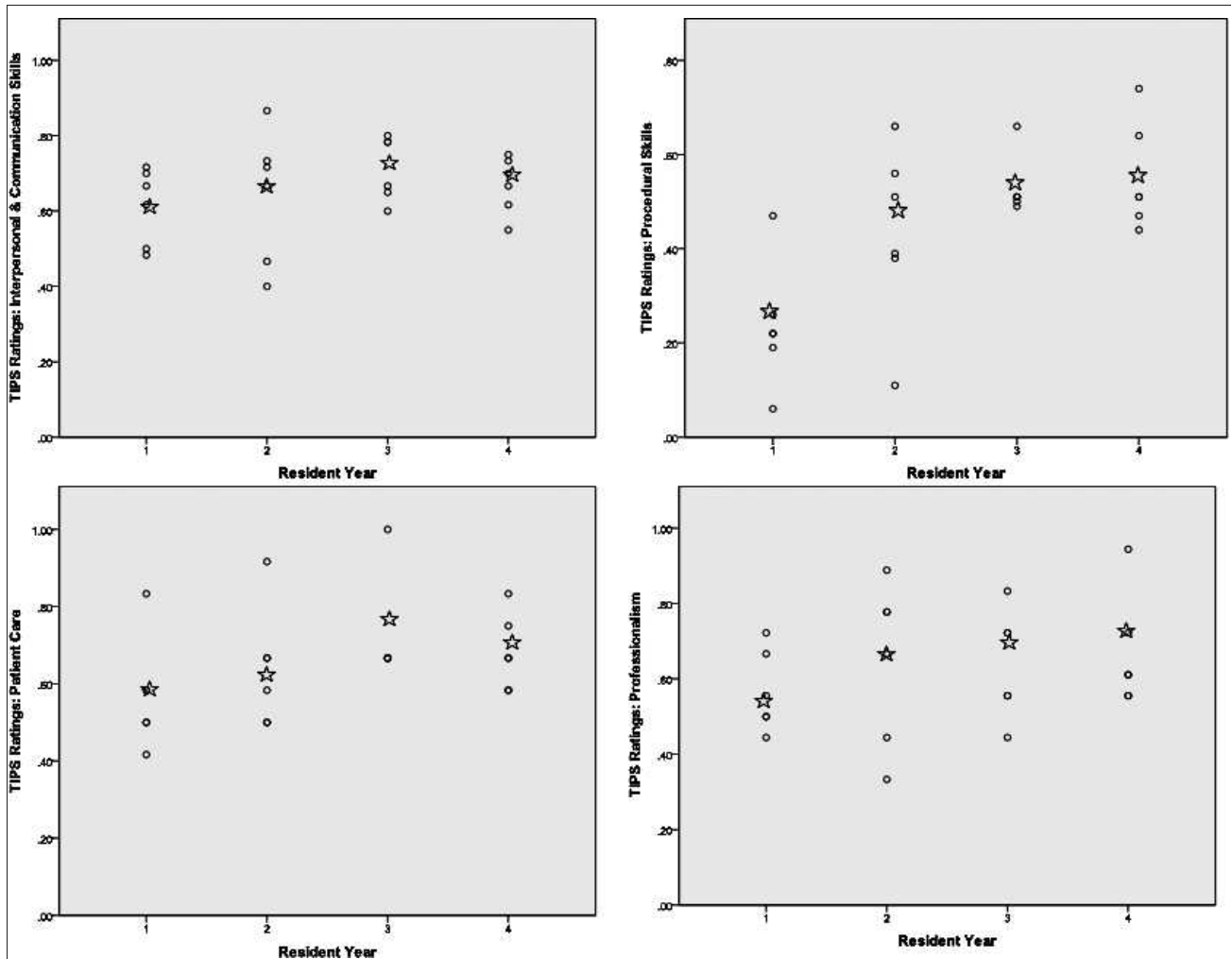


FIGURE 2 | INDIVIDUAL RESIDENT PERFORMANCE ON THE TEST OF INTEGRATED PROFESSIONAL SKILLS (TIPS)

Graphic representation of individual performance on TIPS examination allows identification of poor performers. Outliers compared to class mean (indicated by the star) may be identified, as seen with postgraduate year (PGY)-1 and PGY-2 residents in procedural skills.

Results

Twenty-four residents, postgraduate year (PGY)-1 through PGY-4, participated in the TIPS examination. Mean scores across ACGME competencies appeared to improve with level of training (FIGURE 1), although the difference between PGY levels was found to be significant only for procedural skills (1-way analysis of variance $F = 6.92$, $P < .001$ with Bonferroni post hoc pairwise comparisons). Large variations were observed between residents, as illustrated by the wide standard deviation overall and for each class mean. Residents' mean scores improved between PGY-1 and PGY-4 in all areas, but there was a smaller increase—and sometimes a decrease—in scores between PGY-3 and PGY-4 residents.

Individual residents' scores were examined by using a scatterplot (FIGURE 2) to estimate benchmarks for performance at each level of training. Several outliers were identified visually. The videotapes, checklists,

and comments in the case of outliers were reviewed and discussed with the resident. Faculty completed 658 evaluations for residents. The number of evaluations per resident varied from 15 to 32, with a mean of 21. The mean scores (0–5 Likert scale) for residents for all competencies clustered between 3.5 and 4.3 with little variation (TABLE 2). We found a strong correlation between faculty ratings for medical knowledge and their ratings for every other competency for a given resident on the rotation evaluations (range $r_s = 0.68$ – 0.80 , $P < .001$).

In the competencies of patient care and procedural skills, professionalism, and interpersonal and communication skills, mean TIPS examination scores did not correlate with mean scores for the same competencies on the faculty rotation evaluations (TABLE 3). Outliers identified by the TIPS examination were not identified by the faculty rotation evaluations in either the numerical evaluation or

TABLE 2 FACULTY RATINGS OF RESIDENTS' CLINICAL PERFORMANCE (ROTATIONS)^a

ACGME Competency	PGY-1 Mean (n = 6)	PGY-2 Mean (n = 6)	PGY-3 Mean (n = 6)	PGY-4 Mean (n = 6)	Significance (F)	Correlation with Medical Knowledge Rating
Medical knowledge	3.5 (± 0.36)	3.9 (± 0.32)	3.8 (± 0.27)	3.7 (± 0.36)	F = 2.12, P = .13	Not applicable
Patient care	3.5 (± 0.33)	4.0 (± 0.30)	3.9 (± 0.30)	4.1 (± 0.24)	F = 5.03, P = .01, 1 < 2, 3, 4	0.68 (P < .001)
Interpersonal and communication skills	3.7 (± 0.34)	4.3 (± 0.28)	4.0 (± 0.35)	4.0 (± 0.21)	F = 4.282, P = .02, 1 < 2	0.79 (P < .001)
Professionalism	3.8 (± 0.32)	4.3 (± 0.21)	3.9 (± 0.30)	3.9 (± 0.31)	F = 4.431, P = .02, 1, 4 < 2	0.67 (P < .001)
Systems-based practice	3.5 (± 0.28)	3.9 (± 0.09)	3.7 (± 0.29)	3.6 (± 0.33)	F = 2.61, P = .08	0.73 (P < .001)
Practice-based learning and improvement	3.6 (± 0.27)	3.7 (± 0.28)	3.8 (± 0.28)	3.7 (± 0.39)	F = .48, P = .72	0.73 (P < .001)

Abbreviations: ACGME, Accreditation Council for Graduate Medical Education; PGY, postgraduate year.

^a Mean ratings for performance entered by faculty using the E*Value system, entered on a 0 to 5 Likert scale, according to ACGME competency ratings.

comments sections. A few residents (3 of 24 in each competency) performed poorly on the TIPS (> 1 SD below the mean), but received average or above-average faculty ratings in the same competency. All 24 residents participating in the TIPS examination completed the program evaluation survey. Most comments were positive; residents reported that the intervention was a “very, very useful process” and that it was “wonderful to get real-time feedback in arenas/situations where we don’t often get feedback.” Some negative comments were “OSCEs always feel unnatural,” and that the timing of skills and technical issues with the models was difficult to manage.

Five of 7 faculty preceptors (71%) completed the questionnaire. All responding faculty “strongly agreed” that the TIPS examination helped residents identify strengths/weaknesses, and that they wanted to experience it again as faculty. They agreed that the TIPS examination “taught residents something new” and provided “valuable feedback” and “new information about residents’ performance level.”

Discussion

The TIPS may be valuable in assessing integration of skills and providing an important forum for direct observation

and timely feedback based on performance. Below-average performers on the TIPS examination were not identified through the existing methods of faculty evaluations. The TIPS format highlights difficult-to-teach domains such as interpersonal skills and professionalism. Traditional faculty evaluations were found to have very little variation between residents, and suspiciously high correlation with a given resident’s rating for medical knowledge was attributed to a “halo effect.” TIPS scores generated by standardized patients trained as raters using behaviorally anchored checklists provided different information about the residents’ performance than that obtained through faculty rotation evaluations.

The TIPS examination is a simple model that allows a clinical case to be paired with a focused skill, and it can yield a broad range of feedback about resident performance. This may facilitate the important process of self-directed assessment-seeking that is important to lifelong improvement in physicians. The overall performance of the cohort is useful for program evaluation, and scenarios involving Next Accreditation System Milestones may help make determinations of competency.

There are some limitations to the TIPS examination format. Because of practical constraints, some competencies

TABLE 3 CORRELATION BETWEEN TEST OF INTEGRATED PROFESSIONAL SKILLS EXAMINATION SCORES AND FACULTY ROTATION RATINGS

	Patient Care	Procedural Skills	Professionalism	Interpersonal and Communication Skills
Correlation (Spearman)	0.45 (P = .03)	0.086 (P = .69)	0.36 (P = .09)	0.081 (P = .71)

are only assessed in a single station (systems-based practice and practice-based learning), which does not allow for multiple data points. The selection of 5 scenarios represents a snapshot, but not a complete picture of a resident's ability. Medical knowledge is deliberately not evaluated in TIPS. Instead, residents demonstrate knowledge in the form of transmitted knowledge in patient education (interpersonal and communication skills) or determination of treatment plans (patient care). OSCEs tend to focus more on other competencies, which balance the traditional emphasis on medical knowledge assessment, which can overshadow other competencies. The finding that faculty ratings of other competencies correlated strongly with ratings of medical knowledge further demonstrates this tendency. In this regard, the TIPS examination is providing new assessment information about the performance of these residents.

With a sample of 24, the study is only powered to detect fairly large effects and does not permit exploration of PGY differences or examining rotation evaluations on a faculty or rotation level. The faculty members who scored the procedural elements were not blinded to resident level. How best to rate procedural skills and whether faculty or nonfaculty raters should be used are questions that must be explored.²⁰ Future research will aim to gather additional validity evidence for the TIPS examination by administering it to true novices, students who have not had residency training, and more expert performers who have completed residency training.²⁹ Finally, generalizability of our experience should be assessed by replication in other programs.

Conclusion

The TIPS is a performance-based examination that allows for direct observation of resident interaction with standardized patients, technical skill performance, and searching the medical literature. TIPS examination performance improves with level of training, but variations between individual residents were identified. Feedback from faculty allows timely and individualized learning directed to observed gaps.

References

- Huddle TS, Heudebert GR. Taking apart the art: the risk of anatomizing clinical competence. *Acad Med*. 2007;82(6):536–541.
- Swing SR; International CBME Collaborators. Perspectives on competency-based medical education from the learning sciences. *Med Teach*. 2010;32(8):663–668.
- Epstein RM. Assessment in medical education. *N Engl J Med*. 2007;356(4):387–396.
- Clilliers FJ, Schuwirth LWT, Herman N, Alendorff HJ, van der Vleuten A. A model of the pre-assessment learning effects of summative assessment in medical education. *Adv Health Sci Educ Theory Pract*. 2012;17(1):39–53.
- Shepard LA. The role of assessment in a learning culture. *Educ Researcher*. 2000;29(7):4–14.
- Wagner D, Lypson ML. Centralized assessment in graduate medical education: cents and sensibilities. *J Grad Med Educ*. 2009;1(1):21–27.
- Schuwirth LW, van der Vleuten CP. General overview of the theories used in assessment: AMEE Guide No. 57. *Med Teach*. 2011;33(10):783–797.
- Schuwirth LW, van der Vleuten CPM. Programmatic assessment: from assessment of learning to assessment for learning. *Med Teach*. 2011;33(6):478–485.
- Nasca TJ, Philibert I, Brigham T, Flynn TC. The next GME accreditation system—rationale and benefits. *N Engl J Med*. 2012;366(11):1051–1056.
- Short MW, Jorgenson JE, Edwards JA, Blankenship RB, Roth BJ. Assessing intern core competencies with an objective structured clinical examination. *J Grad Med Educ*. 2009;1(1):30–36.
- Houck WA, Soares-Welch CV, Montori VM, Li JT. Learning the thyroid examination—a multimodality intervention for internal medicine residents. *Teach Learn Med*. 2002;14(1):24–28.
- Lypson ML, Frohna JG, Gruppen LD, Woolscroft JO. Assessing residents' competencies at baseline: identifying the gaps. *Acad Med*. 2009;79(6):564–570.
- Hochberg MS, Kalet A, Zabar S, Kachur E, Gillespie C, Berman RS. Can professionalism be taught: encouraging evidence. *Am J Surg*. 2010;199(1):86–93.
- van der Vleuten CP, Schuwirth LW, Driessen EW, Dijkstra J, Tigelaar D, Baartman LK, et al. A model for programmatic assessment: fit for purpose. *Med Teach*. 2012;34(3):205–214.
- Swing SR, Clyman SG, Holmboe ES, Williams RG. Advancing resident assessment in graduate medical education. *J Grad Med Educ*. 2009;1(2):278–286.
- Franzese CB. Pilot study of an Objective Structured Clinical Examination ("the Six Pack") for evaluating clinical competencies. *Otolaryngol Head Neck Surg*. 2008;138(2):143–148.
- Tudivier F, Rose D, Banks B, Pfortmiller D. Reliability and validity testing of an evidence-based medicine OSCE station. *Fam Med*. 2009;41(2):89–91.
- Varkey P, Natt N, Lesnick T, Downing S, Yudkowsky R. Validity evidence for an OSCE to assess competency in systems-based practice and practice-based learning and improvement: a preliminary investigation. *Acad Med*. 2008;83(8):775–780.
- McGaghie WC, Issenberg SB, Petrusa ER, Scalese RJ. A critical review of simulation-based medical education research: 2003–2009. *Med Educ*. 2010;44(1):50–63.
- Rooney DM, Santos BF, Hungness ES. Fundamentals of laparoscopic surgery (FLS) manual skills assessment: surgeon vs nonsurgeon raters. *J Surg Educ*. 2012;69(5):588–592.
- Sanfey H, Dunnington G. Verification of proficiency: a prerequisite for clinical experience. *Surg Clin North Am*. 2010;90(3):559–567.
- Kim J, Neillpovitz D, Cardinal P, Chiu M. A comparison of global rating scale and checklist scores in the validation of an evaluation tool to assess performance in the resuscitation of critically ill patients during simulated emergencies (abbreviated as "CRM simulator study IB"). *Simul Healthc*. 2009;4(1):6–16.
- Vassiliou MC, Dunkin BJ, Marks JM, Fried GM. FLS and FES: comprehensive models of training and assessment. *Surg Clin North Am*. 2010;90(3):535–558.
- Winkel AF, Lerner V, Zabar SR, Szyld D. A simple framework for assessing technical skills in a resident observed structured clinical examination (OSCE): vaginal laceration repair. *J Surg Educ*. 2013;70(1):10–14.
- Chander B, Kule R, Baiocco P, Chokhavatia S, Kotler D, Poles M, et al. Teaching the competencies: using objective structured clinical encounters for gastroenterology fellows. *Clin Gastroenterol Hepatol*. 2009;7(5):509–514.
- Alevi D, Baiocco PJ, Chokhavatia S, Kotler DP, Poles M, Zabar S, et al. Teaching the competencies: using observed structured clinical examinations for faculty development. *Am J Gastroenterol*. 2010;105(5):973–977.
- Shah B, Miller R, Poles M, Zabar S, Gillespie C, Weinshel E, et al. Informed consent in the older adult: OSCEs for assessing fellows' ACGME and geriatric gastroenterology competencies. *Am J Gastroenterol*. 2011;106(9):1575–1579.
- Zabar S, Hanley K, Stevens DL, Ciotoli C, Hsieh A, Griesser C, et al. Can interactive skills-based seminars with standardized patients enhance clinicians' prevention skills: measuring the impact of a CME program. *Patient Educ Couns*. 2010;80(2):248–252.
- 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.