

Use of the Quality Improvement (QI) Knowledge Application Tool in Assessing Pediatric Resident QI Education

ERIC W. GLISSMEYER, MD
SONJA I. ZINIEL, PhD
JAMES MOSES, MD, MPH

Abstract

Background Assessing the effectiveness of quality improvement curricula is important to improving this area of resident education.

Objective To assess the ability of the Quality Improvement Knowledge Application Tool (QIKAT) to differentiate between residents who were provided instruction in QI and those who were not, when scored by individuals not involved in designing the QIKAT, its scoring rubric, or QI curriculum instruction.

Methods The QIKAT and a 9-item self-assessment of QI proficiency were administered to an intervention and a control group. The intervention was a longitudinal curriculum consisting of 8 hours of didactic QI training and 6 workshops providing just-in-time training for resident QI projects. Two uninformed faculty scored the QIKAT.

Results A total of 33 residents in the intervention group and 27 in the control group completed the baseline and

postcurriculum QIKAT and self-assessment. QIKAT mean intervention group scores were significantly higher than mean control group scores postcurriculum ($P < .001$). Absolute QIKAT differences were small (of 15 points, intervention group improved from a mean score of 12.8 to 13.2). Interrater agreement as measured by kappa test was low (0.09). Baseline self-assessment showed no differences, and after instruction, the intervention group felt more proficient in QI knowledge than controls in 4 of 9 domains tested.

Conclusions The QIKAT detected a statistically significant improvement postintervention, but the absolute differences were small. Self-reported gain in QI knowledge and proficiency agreed with the results of the QIKAT. However, QIKAT limitations include poor interrater agreement and a scoring rubric that lacks specificity. Programs considering using QIKAT to assess curricula should understand these limitations.

Editor's Note: The online version of this article contains Quality Improvement Knowledge Application scenarios adapted to pediatrics and a self-assessment tool to measure quality improvement proficiency.

Eric W. Glissmeyer, MD, is Fellow, Pediatric Emergency Medicine, University of Utah, and Fellow, Intermountain Healthcare Institute for Health Care Delivery Research; **Sonja I. Ziniel, PhD**, is Senior Survey Methodologist, Program for Patient Safety and Quality, Faculty Member, Division of Adolescent and Young Adult Medicine, Department of Medicine, Boston Children's Hospital, and Instructor in Pediatrics, Department of Pediatrics, Harvard Medical School; and **James Moses, MD, MPH**, is Pediatric Director of Quality and Patient Safety, Department of Pediatrics, Boston Medical Center, Associate Program Director, Boston Combined Residency Program in Pediatrics, and Assistant Professor, Department of Pediatrics, Boston University School of Medicine.

Funding: Resident quality improvement efforts were supported by the Fred Lovejoy Research and Education Fund of the Boston Combined Residency Program and a grant from the Program for Patient Safety and Quality, Children's Hospital Boston.

Conflict of interest: The authors declare they have no competing interests.

The authors would like to thank the residents and faculty of the Boston Combined Residency Program in Pediatrics; and Bob Vinci, MD, and Ted Sectish, MD, for their support of the quality improvement curriculum and resident projects.

Corresponding author: Eric W. Glissmeyer, MD, University of Utah Department of Pediatrics, Division of Pediatric Emergency Medicine, 295 Chipeta Way, Salt Lake City, UT 84108, eric.glissmeyer@hsc.utah.edu

Received June 13, 2013; revision received November 18, 2013; accepted January 12, 2014.

DOI: <http://dx.doi.org/10.4300/JGME-D-13-00221>

Introduction

Training in quality improvement (QI) and patient safety is a critical skill for physicians,¹ yet the optimal approach for educating residents in QI and patient safety has not been fully identified.² Wide variation in instructional methods exists among programs and across medical disciplines.² In pediatric training specifically, a recent survey of program directors found a high degree of variation in curricular elements and minimal formal evaluation of trainee QI knowledge, skills, and attitudes.³

Best practices for resident QI education include combining didactic and experiential learning into a longitudinal curriculum with leadership of local experts successful in QI within their clinical practice.⁴⁻⁷ However, few instruments exist that objectively assess gains in QI skills, knowledge, and behavior postinstruction.⁸⁻¹⁰ A tool with evidence of preliminary validity was published during this study but has not been evaluated further.¹¹

The Quality Improvement Knowledge Application Tool (QIKAT) has been used to evaluate QI knowledge acquisition after curricular instruction.¹²⁻¹⁵ The QIKAT assesses differences in resident QI knowledge after curricular instruction, with acceptable to good interrater

agreement when scored by the creators of the tool,¹³ QI instructors,¹⁴ and creators of the scoring rubric.¹⁵ Although the QIKAT has evidence of validity in content only, residency programs may be using the QIKAT to assess QI curricular effectiveness.¹⁵ Important to this use, it is not known how the QIKAT performs when scored by individuals not involved in design of the tool, curricular instruction, and scoring rubric.

The objective of our study was to assess the ability of the QIKAT to differentiate between residents who were provided with instruction in QI and those who were not, when scored by individuals not involved in the QI curriculum delivery, the development of the QIKAT, or the scoring rubric.

Methods

Our study used a pre-post quasi-experimental test design. The intervention was a longitudinal QI curriculum delivered to second-year pediatrics residents in the 2010–2011 academic year. Third-year pediatrics residents not exposed to the longitudinal QI curriculum formed the concurrent control group. Baseline measurements included the QIKAT and a QI self-assessment questionnaire of QI proficiency. After intervention group curricular instruction, both groups completed 3 QIKAT scenarios and the self-assessment.

Assessment Tools

The QIKAT presents the learner with clinical scenarios that fail in at least 1 Institute of Medicine dimension of health care performance.¹⁶ We modified the clinical content of the QIKAT scenarios to include disease states familiar to pediatrics residents (provided as online supplemental material). After reading the scenario, the learner responds to 3 free-text questions asking about an aim for improvement, measures for assessment, and a change proposal that could be tested.

The 6 pediatric QIKAT scenarios underwent beta testing for content clarity by third-year pediatrics residents not otherwise involved. The structures of the 3 baseline and 3 postintervention QIKAT scenarios were similar, but their clinical context differed to avoid direct recall of responses. We measured resident self-perception of QI skills with a QI knowledge self-assessment tool¹⁷ (provided as online supplemental material), which has been used in other studies to demonstrate self-perception of QI knowledge gained after curricular instruction.^{13,15} Residents rated their proficiency in 9 QI skills using a Likert scale from 1 (not proficient) to 5 (very proficient). The QIKAT scenarios adapted to pediatrics and the self-assessment tool are available as online supplemental material.

QI Curriculum Description

We designed our QI curriculum using methodology suggested by experts.^{4,18} It incorporated guiding principles

What was known

Assessing the effectiveness of a quality improvement (QI) curriculum is important to improving QI education and, through this, skills acquisition.

What is new

A test of the ability of the QI Knowledge Application Tool (QIKAT) to differentiate between residents who were provided QI instruction and those without instruction, when scored by faculty not involved in designing the QIKAT, its scoring rubric, or QI curriculum instruction.

Limitations

Limited sample, poor interrater agreement, and use of only 1 scoring rubric for scoring QIKAT responses.

Bottom line

Programs considering using the QIKAT to assess curricular effectiveness should understand the limitations of the tool and its scoring rubric.

for teaching others to lead change by combining didactic and experiential learning with leadership from clinicians demonstrating continuous improvement in their own work.⁶ The intervention group received the QI curriculum longitudinally through the 2010–2011 academic year. Faculty trained in QI methodology and health services research taught 8 noon conference didactic sessions. Six 3-hour workshops provided just-in-time training and focused on QI skills matched to the needs of residents' projects.

QI projects were designed and carried out by residents working in teams and mentored by faculty. Projects included (1) reducing time to lumbar puncture in febrile infants in the emergency department; (2) enhancing parental understanding of discharge instructions by standardizing discharge information; (3) increasing primary care follow-up and controller medication use in patients discharged after asthma exacerbation; (4) increasing the availability of working mobile computer workstations during intensive care unit rounds; (5) increasing referral of primary care patients with obesity to a multidisciplinary team; and (6) incorporation of postpartum depression screening at 2-, 4-, and 6-month well-child visits. Residents in the intervention group were required to propose a QI project, present interim data, conduct at least 2 plan-do-study-act cycles (many undertook several), and present final data and outcomes.

Scoring of QIKAT Responses

Two raters scored responses to the open-ended questions of the QIKAT scenarios. Both of the raters had graduated fellowships in health care delivery research and held faculty positions in QI in their respective institutions. We used a scoring rubric, described previously,¹⁵ which awards a maximum of 5 points per scenario, 3 scenarios per test, for

a maximum of 15 points. Maximum points per subsection were 2 points for the aim, 1 point for measures, 1 point for change proposal, and 1 point for answers that were related to one another. Raters were blinded to the study objectives and had no role in the QI curriculum. After scoring a random sample of 20 QIKAT responses, raters met to discuss scoring decisions. Differences of more than 1 point per 15-point QIKAT were resolved; thereafter, the raters scored responses independently.

The Institutional Review Board of Boston Children’s Hospital approved this study and granted a waiver of informed consent.

Analyses

All analyses were performed using Stata version 12.1 software (Stata Corp LP, College Station, TX). Analyses consisted of 2 parts: the first part focused on the influence of the QI curriculum over time and differences between the comparison and intervention groups, and the second focused on the scores of the individual raters and their agreement. The significance of mean score comparisons between intervention and control groups was assessed using the 2-sample Wilcoxon rank sum (Mann-Whitney *U*) test due to the nonnormal distribution of scores. Wilcoxon matched pairs signed rank test was used to test mean score differences between baseline and postcurriculum measurements. We also used linear regression analyses to assess the unique influence of having completed the QI curriculum on postcurriculum scores by controlling for baseline scores for each of the QIKAT subsections. Predictor variables were cubed to normalize the regression residuals to meet regression analysis assumptions. Scoring differences between the 2 raters were evaluated through Wilcoxon rank sum tests and correlations (tetrachoric or Spearman rank correlation coefficients). Interrater agreement was measured through Cohen’s kappa test. For all analyses, a *P* value of < .05 was considered statistically significant.

Results

The intervention group included 36 residents and the control group 27 residents. Three residents in the intervention group did not complete the postcurriculum self-assessment or QIKAT and were excluded from analysis. None of the residents in the intervention or control group reported having received prior formal training in QI methods.

Overall—QIKAT

Data for QIKAT scores by intervention and control groups are shown in TABLE 1. Mean intervention group scores were significantly higher than mean control group scores postcurriculum (*P* < .001). Baseline mean QIKAT scores for the intervention and control groups were similar (*P* = .06). Within the intervention group, postcurriculum

COMPARISON OF BASELINE AND POSTCURRICULUM QIKAT SCORES FOR INTERVENTION AND CONTROL GROUPS											
TABLE 1		Maximum Score Possible	Baseline Mean (±SE)	Postcurriculum Mean (±SE)	Comparison Baseline to Postcurriculum P Value	Baseline Control vs Intervention P Value	Postcurriculum Control vs Intervention P Value				
Overall	Control (n = 27)	15	11.8 (0.45)	11.6 (0.35)	.43	.06	< .001				
	Intervention (n = 33)	15	12.8 (0.29)	13.2 (0.30)	.44						
Aim	Control (n = 27)	6	4.5 (0.22)	4.2 (0.18)	.26	.37	.01				
	Intervention (n = 33)	6	4.8 (0.15)	4.8 (0.16)	.92						
Measures	Control (n = 27)	3	2.9 (0.08)	2.8 (0.07)	.95	.23	.81				
	Intervention (n = 33)	3	2.8 (0.06)	2.9 (0.07)	.07						
Change proposal	Control (n = 27)	3	2.2 (0.16)	2.4 (0.15)	.37	.07	.04				
	Intervention (n = 33)	3	2.7 (0.07)	2.8 (0.05)	.07						
Relatedness	Control (n = 27)	3	2.2 (0.15)	2.2 (0.14)	.49	.06	.002				
	Intervention (n = 33)	3	2.6 (0.08)	2.7 (0.06)	.54						

TABLE 2 COMPARISON BY RATER OF BASELINE AND POSTCURRICULUM QIKAT SCORES FOR INTERVENTION AND CONTROL GROUP

			Maximum Score Possible	Baseline Mean ($\pm$ SE)	Postcurriculum Mean ($\pm$ SE)	Comparison Baseline to Postcurriculum <i>P</i> Value	Baseline Control vs Intervention <i>P</i> Value	Postcurriculum Control vs Intervention <i>P</i> Value
Overall	Rater 1	Control (n = 27)	15	11.5 (0.46)	11.8 (0.44)	.40	.03	.05
		Intervention (n = 33)	15	12.8 (0.31)	12.8 (0.35)	.82		
	Rater 2	Control (n = 27)	15	12.1 (0.49)	11.4 (0.39)	.09	.25	< .001
		Intervention (n = 33)	15	12.9 (0.35)	13.5 (0.29)	.30		
Aim	Rater 1	Control (n = 27)	6	4.3 (0.25)	4.3 (0.24)	.91	.38	.23
		Intervention (n = 33)	6	4.6 (0.17)	4.6 (0.19)	.87		
	Rater 2	Control (n = 27)	6	4.6 (0.23)	4.1 (0.21)	< .05	.34	.002
		Intervention (n = 33)	6	4.9 (0.17)	5.0 (0.18)	.71		
Measures	Rater 1	Control (n = 27)	3	2.8 (0.11)	2.8 (0.08)	.73	.44	.53
		Intervention (n = 33)	3	2.7 (0.09)	2.9 (0.08)	.17		
	Rater 2	Control (n = 27)	3	2.9 (0.05)	2.9 (0.07)	.32	.36	.53
		Intervention (n = 33)	3	2.9 (0.06)	2.9 (0.07)	.52		
Change proposal	Rater 1	Control (n = 27)	3	2.1 (0.18)	2.4 (0.16)	.20	.01	.10
		Intervention (n = 33)	3	2.6 (0.10)	2.7 (0.09)	.55		
	Rater 2	Control (n = 27)	3	2.4 (0.16)	2.4 (0.15)	.82	.19	.002
		Intervention (n = 33)	3	2.7 (0.09)	2.9 (0.04)	.03		
Relatedness	Rater 1	Control (n = 27)	3	2.3 (0.13)	2.3 (0.14)	.87	.004	.02
		Intervention (n = 33)	3	2.8 (0.07)	2.7 (0.09)	.74		
	Rater 2	Control (n = 27)	3	2.1 (0.19)	2.0 (0.18)	.56	.32	.003
		Intervention (n = 33)	3	2.4 (0.12)	2.7 (0.08)	.16		

TABLE 3 OVERALL AND SUBSECTION QIKAT SCORES BY RATER^a

		Mean ($\pm$ SE)	P Value	Correlation	P Value	Kappa	P Value
Overall	Rater 1	24.6 (0.42)	.17	0.66	< .001	0.09	.01
	Rater 2	25.1 (0.45)					
Aim	Rater 1	8.9 (0.20)	.01	0.63	< .001	0.05	.16
	Rater 2	9.4 (0.23)					
Measures	Rater 1	5.7 (0.09)	.07	0.56	< .001	0.51	< .001
	Rater 2	5.8 (0.07)					
Change proposal	Rater 1	5.0 (0.15)	< .005	0.57	< .001	0.33	< .001
	Rater 2	5.3 (0.13)					
Relatedness	Rater 1	5.1 (0.12)	< .001	0.66	< .001	0.03	.37
	Rater 2	4.6 (0.17)					

^a Scores are summations of baseline and postcurriculum Quality Improvement Knowledge Application Tool (QIKAT).

mean QIKAT scores were not significantly higher than baseline QIKAT scores ($P = .44$). However, multivariate regression analysis indicated a significant increase in postcurriculum scores for the intervention group compared to the control group, accounting for baseline scores ($P = .007$; detailed results not shown).

Subsections—QIKAT

For all subsections, distributions of baseline and postcurriculum scores were left-skewed. There were no statistically significant baseline differences between control and intervention groups (TABLE 1). Postcurriculum, scores for 3 of the 4 subsections (aim, change proposal, and relatedness) were significantly higher for the intervention group ($P = .009$; $P = .04$; $P = .002$). Changes in mean scores between baseline and postcurriculum assessments within each group and subsections were not significant in bivariate analyses. Multivariate regression analysis confirmed this finding for the aim, measures, and change proposal subsections but showed a significant increase in scores for the relatedness subsection for the intervention group ($P < .001$; detailed results not shown).

QIKAT Scores by Rater

Analyses of QIKAT scores by rater are shown in TABLE 2. Both of the raters identified a postcurriculum difference between intervention and control groups in overall scores and the relatedness subsection. Neither rater identified a difference in measures subsection scores between control and intervention groups postcurriculum. Rater 2 identified postcurriculum differences between the groups in the aim and change proposal subsections, and rater 1 identified a score difference at baseline between the intervention and control groups and for the change proposal and relatedness subsections.

TABLE 3 shows interrater agreement between raters 1 and 2. As raters were blinded to when each scenario was administered, the analyses combine the ratings of all 6 scenarios. Although the overall scores and scores in the measures subsection did not differ significantly between raters, the scores for all other subsections did. Correlation coefficients for overall QIKAT and the subsections between the 2 raters were fairly high and statistically significant, but the value of kappa was generally low.

Self-Assessment Results

No baseline differences between the intervention and control groups were identified using the self-assessment tool (TABLE 4). Postcurriculum, the intervention group's self-assessment ratings were generally higher than the ratings of the control group, reaching significance in 4 of 9 QI skills. Within the intervention group, self-assessment scores significantly increased after QI instruction for all 9 QI skills. The control group self-assessment scores significantly increased only for ability to "identify a quality problem related to patient care."

Discussion

Our evaluation of the QIKAT in a quasi-experimental setting with a concurrent control group found that the QIKAT successfully distinguished intervention from control group after instruction in QI. However, we identified challenges to the performance of the QIKAT. Baseline scores were high, and absolute score differences for the intervention group were small, with a mean increase in QIKAT score of 0.34 of 15 total points (12.8 to 13.2). Also, interrater agreement as measured by kappa test was poor (0.09). Combined baseline and postcurriculum QIKAT

TABLE 4 RESULTS FROM SELF-ASSESSMENT OF QI PROFICIENCY

Item	Control Group			Intervention Group			Baseline Control vs Intervention Group	Postcurriculum Control vs Intervention Group
	Before: Mean ($\pm$ SE)	After: Mean ($\pm$ SE)	P Value	Before: Mean ($\pm$ SE)	After: Mean ($\pm$ SE)	P Value	P Value	P Value
Identify a quality problem related to patient care	3.3 (0.15)	3.8 (0.13)	.02	3.6 (0.14)	3.9 (0.12)	.02	.17	.35
Develop and focus an aim related to a quality problem	2.8 (0.20)	3.2 (0.17)	.07	2.9 (0.17)	3.6 (0.12)	< .001	.74	.02
Identify outcome and process measures appropriate for a clinical problem	2.7 (0.12)	3.1 (0.18)	.10	2.6 (0.15)	3.4 (0.15)	< .001	.43	.25
Identify changes in practice to improve processes and outcomes of care	2.9 (0.17)	3.2 (0.17)	.28	2.9 (0.17)	3.6 (0.15)	< .001	.84	.02
Demonstrate how to use several cycles of change to improve care delivery	2.1 (0.15)	2.4 (0.17)	.09	2.0 (0.15)	3.2 (0.17)	< .001	.63	.004
Formulate a data plan related to demonstrating a change results in improvement	2.2 (0.18)	2.5 (0.16)	.13	2.1 (0.15)	2.9 (0.17)	< .001	.83	.09
Use run/control charts in displaying results of change in an effective manner	1.7 (0.18)	1.9 (0.15)	.34	1.4 (0.10)	2.3 (0.18)	< .001	.55	.12
Create an interdisciplinary improvement team and assign roles necessary for improvement success	1.9 (0.19)	2.2 (0.16)	.09	2.2 (0.16)	3.1 (0.17)	< .001	.16	< .001
Ensure change tested is implemented into practice and sustained	2.0 (0.17)	2.2 (0.15)	.30	2.0 (0.15)	2.56 (0.15)	< .001	.89	.15

Abbreviation: QI, quality improvement.

TABLE 5 COMPARISON OF STUDIES USING THE QIKAT				
Comparative Analysis of QIKAT Studies				
	This Study	Ogrinc et al ¹³ 2004	Varkey et al ¹⁴ 2006	Vinci et al ¹⁵ 2010
No. in experiment/no. in control	36/27	11/22	7/0	36/36
Curriculum length (months)	12	1	1	2 ^a
Study aim	Test QIKAT performance	Describe QI curriculum	Describe QI curriculum	Assess curriculum effectiveness
Experimental design	Pre-post	Pre-post	Pre-post	Historical control
Experimental group	PL-2 residents	PL-2–PL-4 residents	Residents and nursing graduate students	PL-2 residents
Nature of participation	Curriculum requirement	Volunteer participants	Curriculum requirement	Curriculum requirement
Number of QIKAT scorers	2	2	1	3
QIKAT scoring by designers of scoring rubric	No	Yes	No	Yes
QIKAT scoring by curricular instructors	No	Yes	Yes	Yes
QIKAT scoring by authors	No	Yes	Yes	Yes
Kappa determination	Overall	Overall	None	Random sample of 10 responses
Kappa	0.08	0.2–0.42	NA	0.8

Abbreviations: QIKAT, Quality Improvement Knowledge Application Tool; PL, pediatrics level; NA, not applicable.
^aQuality improvement instruction occurred in the second month, and postcurricular QIKAT testing occurred immediately thereafter.

scores from raters 1 and 2 were significantly different in 3 of 4 QIKAT subsections.

Our results highlight important limitations in the QIKAT as an assessment tool. The QIKAT has limited evidence of validity as a tool to assess QI knowledge gained after QI instruction. Similar to previous studies, we found the QIKAT distinguished intervention from control groups after QI instruction. Major differences in our study were lower absolute differences in QIKAT scores and lower kappa results. In previous studies, the QIKAT was used either to describe the curriculum^{13,14} or to assess curricular effectiveness¹⁵ and did not include baseline and postcurriculum control groups.^{13–15} The absence of a concurrent control group may not permit causal inferences of the ability of the QIKAT to assess knowledge gained from curricular instruction.¹⁹ TABLE 5 compares the design of this study to previous uses of the QIKAT. Finally, the QIKAT does not test knowledge of other elements for successful system improvement such as skills for functioning in multidisciplinary teams and project prioritization and use of tools for outcome measurement.^{5,6,12} The scoring rubric used in this study appeared to lack specificity. Raters were asked to make qualitative judgments (ie, good, excellent). Few points were available, and most scores had a

binary value. We believe this led to “grade inflation.” Improvements in the QIKAT could include more difficult questions, testing other QI knowledge areas, and an enhanced scoring rubric. Many of these proposed improvements to the QIKAT are features of another QI knowledge assessment tool, the Systems Quality Improvement Training and Assessment Tool, published after completion of this study.¹¹

Limitations of our study include the poor interrater agreement as measured by kappa test. However, although absolute agreement between raters may be poor, correlation coefficients between raters were statistically significant. QIKAT responses were scored using only 1 scoring rubric, and we do not know how results would have been different if another rubric had been used.

Conclusion

The QIKAT distinguishes intervention from control group after curricular instruction in QI. In this study, overall scores were high, and interrater agreement was poor. The QIKAT in its current form lacks specificity and is not generalizable due to limitations in scoring. We caution educators about these limitations when considering use of the QIKAT in curricular assessment.

References

- 1 Accreditation Council for Graduate Medical Education. ACGME program requirements for graduate medical education in pediatrics. 2013. https://www.acgme.org/acgmeweb/Portals/0/PFAssets/2013-PR-FAQ-PIF/320_pediatrics_07012013.pdf. Accessed March 5, 2014.
- 2 Boonyasai RT, Windish DM, Chakraborti C, Feldman LS, Rubin HR, Bass EB. Effectiveness of teaching quality improvement to clinicians: a systematic review. *JAMA*. 2007;298(9):1023–1037.
- 3 Moses JM, Craig MS, Mann KJ. A Survey of Quality Improvement (QI) Educational Practices in Pediatric Residency Programs: Program Director's (PPD) Perspectives. Boston, MA: Pediatric Academic Societies; 2012.
- 4 Institute for Healthcare Improvement. Eight knowledge domains for health professional students. 2014. <http://www.ihl.org/offerings/ihopenschool/resources/Pages/Publications/EightKnowledgeDomainsForHealthProfessionStudents.aspx>. Accessed January 15, 2014.
- 5 Ogrinc G, Headrick LA, Mutha S, Coleman MT, O'Donnell J, Miles PV. A framework for teaching medical students and residents about practice-based learning and improvement, synthesized from a literature review. *Acad Med*. 2003;78(7):748–756.
- 6 Batalden PB, Kerrigan CL. *Lessons Learned in Changing Healthcare ... and How We Learned Them*. Toronto, ON: Longwoods Publishing; 2010.
- 7 Patow CA, Karpovich K, Riesenberger LA, Jaeger J, Rosenfeld JC, Wittenbreer M, et al. Residents' engagement in quality improvement: a systematic review of the literature. *Acad Med*. 2009;84(12):1757–1764.
- 8 Leenstra JL, Beckman TJ, Reed DA, Mundell WC, Thomas KG, Krajicek BJ, et al. Validation of a method for assessing resident physicians' quality improvement proposals. *J Gen Intern Med*. 2007;22(9):1330–1334.
- 9 Wittich CM, Beckman TJ, Drefahl MM, Mandrekar JN, Reed DA, Krajicek BJ, et al. Validation of a method to measure resident doctors' reflections on quality improvement. *Med Educ*. 2010;44(3):248–255.
- 10 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.
- 11 Lawrence RH, Tomolo AM. Development and preliminary evaluation of a practice-based learning and improvement tool for assessing resident competence and guiding curriculum development. *J Grad Med Educ*. 2011;3(1):41–48.
- 12 Wong BM, Levinson W, Shojania KG. Quality improvement in medical education: current state and future directions. *Med Educ*. 2012;46(1):107–119.
- 13 Ogrinc G, Headrick LA, Morrison LJ, Foster T. Teaching and assessing resident competence in practice-based learning and improvement. *J Gen Intern Med*. 2004;19(5, pt 2):496–500.
- 14 Varkey P, Reller MK, Smith A, Ponto J, Osborn M. An experiential interdisciplinary quality improvement education initiative. *Am J Med Quality*. 2006;21(5):317–322.
- 15 Vinci LM, Oyler J, Johnson JK, Arora VM. Effect of a quality improvement curriculum on resident knowledge and skills in improvement. *Qual Saf Health Care*. 2010;19(4):351–354.
- 16 Institute of Medicine; Committee on Quality of Health Care in America. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: National Academy Press; 2001.
- 17 Langley GJ. *The Improvement Guide: A Practical Approach to Enhancing Organizational Performance*. 2nd ed. San Francisco: Jossey-Bass; 2009.
- 18 Ogrinc GS, Headrick L. *Fundamentals of Health Care Improvement: A Guide to Improving Your Patients' Care*. Oak Brook Terrace, IL: Joint Commission Resources; 2008.
- 19 Hadish WR, Cook TD, Campbell DT. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin; 2001.