

# Development and Preliminary Evaluation of a Practice-Based Learning and Improvement Tool for Assessing Resident Competence and Guiding Curriculum Development

RENÉE H. LAWRENCE, PHD  
ANNE M. TOMOLO, MD, MPH

## Abstract

**Background** Although practice-based learning and improvement (PBLI) is now recognized as a fundamental and necessary skill set, we are still in need of tools that yield specific information about gaps in knowledge and application to help nurture the development of quality improvement (QI) skills in physicians in a proficient and proactive manner. We developed a questionnaire and coding system as an assessment tool to evaluate and provide feedback regarding PBLI self-efficacy, knowledge, and application skills for residency programs and related professional requirements.

**Methods** Five nationally recognized QI experts/leaders reviewed and completed our questionnaire. Through an iterative process, a coding system based on identifying key variables needed for ideal responses was developed to score project proposals. The coding system comprised 14 variables related to the QI projects, and an additional 30 variables related to the core knowledge concepts related to PBLI. A total of 86 residents completed the questionnaire, and 2 raters coded their open-ended

responses. Interrater reliability was assessed by percentage agreement and Cohen  $\kappa$  for individual variables and Lin concordance correlation for total scores for knowledge and application. Discriminative validity (*t* test to compare known groups) and coefficient of reproducibility as an indicator of construct validity (item difficulty hierarchy) were also assessed.

**Results** Interrater reliability estimates were good (percentage of agreements, above 90%;  $\kappa$ , above 0.4 for most variables; concordances for total scores were  $R = .88$  for knowledge and  $R = .98$  for application).

**Conclusion** Despite the residents' limited range of experiences in the group with prior PBLI exposure, our tool met our goal of differentiating between the 2 groups in our preliminary analyses. Correcting for chance agreement identified some variables that are potentially problematic. Although additional evaluation is needed, our tool may prove helpful and provide detailed information about trainees' progress and the curriculum.

At the time of this writing, both authors were at the Department of Veterans Affairs Health Services Research and Development's Center for Implementation Practice and Research Support at the Louis Stokes Cleveland Veteran's Affairs Medical Center. Renée H. Lawrence, PhD, is a Health Research Scientist. Anne M. Tomolo, MD, MPH, was a Staff Physician and an Assistant Professor, Department of Medicine at Case Western Reserve University School of Medicine. Dr. Tomolo is currently affiliated with the Atlanta Veterans Affairs Medical Center.

This work was funded in part by a Department of Veterans Affairs' Health Services Research and Development Service Health Services Research & Development grant (small grant award 08-194, Dr. D.C. Aron and Dr. A. M. Tomolo). All statements and descriptions expressed herein do not necessarily reflect the opinions or positions of the Department of Veterans Affairs.

We would like to acknowledge the time and effort contributed by the residents and the faculty and helpful discussions with Brook Watts, Elaine Thallner, David Aron, and Mamta Singh.

Corresponding author: Renée H. Lawrence, PhD, Louis Stokes Cleveland Department of Veterans Affairs Medical Center, 10701 East Boulevard, 111W, Cleveland, OH 44106, 216.791.3800, Renee.Lawrence2@va.gov

Received June 11, 2010; revision received August 24, 2010 and September 20, 2010; accepted October 18, 2010.

DOI: 10.4300/JGME-D-10-00102.1

## Background

The Accreditation Council for Graduate Medical Education (ACGME) released its new framework for physician competence in 1999 and identified Practice-Based Learning and Improvement (PBLI) as one of the 6 basic competencies for physicians.<sup>1,2</sup> Subsequently the American Board of Medical Specialties adopted PBLI as a physician core competency required for maintenance of certification, highlighting the importance for practicing physicians.<sup>2,3</sup>

To gauge PBLI competence, various assessment approaches have emerged.<sup>4-8</sup> Some use strategies that rely on examining changes in existing quality of care measures<sup>9,10</sup> or clinical outcomes,<sup>11,12</sup> whereas others involve assessment instruments that more directly measure knowledge and application skills. Some instruments rely on or include self-assessment of confidence (self-efficacy) and/or practice reflection,<sup>10,13,14</sup> knowledge,<sup>12,15-18</sup> behavior,<sup>9,13</sup> attitudes,<sup>11,14,15,17,19</sup> and/or application skills.<sup>11,12,15-20</sup>

Assessment instruments for use by external raters have emerged and have been used to tap knowledge<sup>15,16,19,21,22</sup> and application skills via proposal development.<sup>23</sup> Additional approaches entail evaluating actual projects or presentations (implementation and sustainability) to demonstrate competency<sup>11,16-18</sup> and assessing learner satisfaction with the curriculum or learning experience.<sup>16,18,22</sup>

Although these approaches helped stimulate thinking about PBLI and have provided a foundation for assessment, many instruments do not address a significant portion of fundamental principles of PBLI, do not provide enough detail of the knowledge attained, and/or blend evaluation of key concepts, making it difficult to identify specific knowledge and skill gaps within the larger content area.<sup>4-6,8</sup> Similarly, although changes to clinical outcomes are an important product of PBLI projects, as an isolated measure of competence these clinical outcomes may not completely represent improvement in knowledge and skills, especially if projects are performed by groups of trainees.

Tools that yield specific information about gaps in knowledge and application are critical to informing the development of trainees, especially if we want to nurture development of physicians who are proficient and proactive in QI.<sup>5</sup> We sought to develop a tool that helps address these issues and can be used for formative and summative evaluations along a broad continuum of learners.<sup>5</sup>

## Method

We developed a Systems Quality Improvement Training and Assessment Tool (SQI TAT) consisting of a questionnaire and a coding system for scoring open-ended responses. The tool addresses the following domains: application skills, self-efficacy (confidence), and knowledge.

### Development of the Questionnaire

The questionnaire contains 2 parts. Part A is the application component and is completed first. Part B contains the efficacy and knowledge components. The items completed by the residents are included as an APPENDIX.

**Part A** The application component was designed to (1) mirror the context of application after residency (identifying a problem, proposing a project to address that problem, convincing management of its potential added value); and (2) assess the ability to develop a project with awareness of continuous quality improvement (CQI) principles. Rather than list the critical sections to include (eg, aims, measures, etc.), we based our format on recall and understanding, rather than recognition via prompting. Instructions for the part A open-ended format are “Based on your clinical experiences, develop a project that would help to improve any aspects of patient care. Please provide enough information so that someone unfamiliar with the context would know what to do, how to do it and why.”

**Part B** To create closed-ended items to assess self-efficacy and comfort corresponding to core considerations that formed a developmental trajectory, we evaluated previous efficacy items and/or scales and relied, in particular, on the work of Djurichich and colleagues<sup>15,17</sup> and Ogrinc et al.<sup>16</sup> Specifically, we felt the Djurichich and colleagues<sup>15,17</sup> efficacy item, “I believe I am able to develop and implement a CQI project,” needed to be split because it appeared to combine 2 steps in a developmental progression (ie, self-efficacy related to developing a project and self-efficacy related to implementing one), and we felt that a few more items were necessary to capture the core consideration in achieving the competency. The Ogrinc et al.<sup>16</sup> 10-item efficacy scale seemed to include items that may be more specific (eg, tool specific) than needed for a general developmental progression. Similarly, some of the items seemed less germane (eg, interdisciplinary collaboration) for a preliminary analysis of developmental progression, especially because we were trying to keep the tool as brief as possible. We agree that the Ogrinc et al.<sup>16</sup> item “formulating a data plan” is a core consideration but felt more information was needed. In addition, we felt self-efficacy related to teaching was important for the developmental trajectory. Accordingly, we used the following 4 closed-ended items (5-point Likert scale): (1) developing a plan, (2) developing a plan that takes contextual considerations and restraints into consideration, (3) implementing a plan, and (4) teaching (TABLE 1).

Basic knowledge was assessed by 6 open-ended items (see bottom of TABLE 2) developed to represent fundamental principles of CQI that complement the application component without redundancy. The items focus on core concepts that need to be understood to adequately develop a project, although the actual terms or concepts may not be defined in a proposal (eg, define common-cause variation). Similar to knowledge items in other tools, we included an item about the Model for Improvement.<sup>15-17</sup> Yet we wanted to assess the broader sense of context associated with the plan-do-study-act cycle that is part of the Model for Improvement including understanding of the 3 guiding questions (eg, What are we trying to accomplish?) and to define (rather than merely list) the elements of the plan-do-study-act cycle (see TABLE 3, where the coding scheme is provided for this section). Uniquely, when compared with other knowledge tools, we included items about change concepts, the cause-effect diagram, and special and common cause variation.

To help establish content and face validity, the questionnaire was reviewed by 4 former National Veterans Affairs Quality Scholars who had participated in the National Veterans Affairs Quality Scholars Fellowship Program at the Cleveland site and were trained in the principles of PBLI. All agreed that the instrument tapped relevant and core aspects of PBLI. In addition, 3 former residents in internal medicine provided feedback that helped us finalize the format.

**TABLE 1** COMPARISON BETWEEN RESIDENTS WHO HAD (N = 23) AND WHO HADN'T (N = 63) PARTICIPATED IN A PRIOR PRACTICE-BASED LEARNING AND IMPROVEMENT (PBLI) COURSE

	No Prior PBLI, Mean (SD)	Prior PBLI, Mean (SD)
Total application score (proposed project)	2.97 (1.91)	3.22 (2.09)
Different domains of self-efficacy and comfort (scale: 1 [strongly disagree] to 5 [strongly agree])		
Project development (I believe I am able to develop a CQI project)	3.28 (0.70)	3.87 (0.76) <sup>a</sup>
Developing data collection plan (I believe I am comfortable developing a data collection plan consistent with time and resource limitations)	3.21 (0.77)	3.74 (0.62) <sup>a</sup>
Project implementation (I believe I am able to implement a CQI project)	3.19 (0.61)	3.65 (0.65) <sup>a</sup>
Teaching (I believe I am able to teach CQI principles)	2.54 (0.89)	2.76 (0.80)
Total knowledge score	0.91 (1.63)	3.09 (2.63) <sup>a</sup>

Abbreviation: CQI, continuous quality improvement.

<sup>a</sup> Difference is significant at  $P < .05$  level.**TABLE 2** SUMMARY OF INTERRATER RELIABILITY: AGREEMENT AND COVARIATION

	No. of Variables Coded	No. of Points (Ideal)	Range of Agreement <sup>a,b</sup> (%)	Lin Concordance <sup>a</sup>	Observed Range <sup>c</sup> (Mean; Mode) (N = 86)
Application sections					
Background (see TABLE 4 for scoring system)	4	5	95–100	...	0–4 (1.34; 1)
Aim (see TABLE 4 for scoring system)	2	4	95–100	...	0–2 (0.07; 0)
Intervention	2	5	95–100	...	0–5 (1.1; 1)
Measurement	2	3	91–100	...	0–3 (0.38; 0)
Impact	1	3	100	...	0–3 (0.14; 0)
Next steps	3	9	100	...	0–3 (0.03; 0)
<b>Total application score</b>	<b>14</b>	<b>29</b>	...	0.88	0–10 (3.03; 2)
Knowledge open-ended items					
Describe what is meant by a change concept	3	7	100	...	0–3 (0.42; 0)
Describe how a cause-effect diagram is created	9	11	95–100	...	0–5 (0.38; 0)
Describe the elements of the improvement model (see TABLE 3 for scoring system)	8	9	86–100	...	0–5 (0.35; 0)
Define common cause (random) variation	3	6	90–100	...	0–3 (0.10; 0)
Define special cause (assignable) variation	3	6	95–100	...	0–5 (0.41; 0)
Why is the distinction between random and special cause variations important?	4	12	95–100	...	0–3 (0.17; 0)
<b>Total knowledge score</b>	<b>30</b>	<b>51</b>	...	0.98	0–10 (1.49; 0)

<sup>a</sup> The final sets of independent coding are reported ( $n = 21$ ).<sup>b</sup> The results for correcting for chance for application variables using Cohen  $\kappa$  were 6 coefficients at 1.0; 2 between 0.75 and 0.99; 1 between 0.4 and 0.74 (indicating fair to good agreement beyond chance); 1 below 0.4 (poor agreement); and 4 were indeterminate (no variation with regard to present/absent because coders agreed 100% that the variable was absent). The results for correcting for chance for knowledge variables using Cohen  $\kappa$  were 8 coefficients at 1.0; 4 between 0.75 and 0.99 (indicating excellent agreement beyond chance); 3 between 0.4 and 0.74 (indicating fair to good agreement beyond chance); 2 below 0.4 (poor agreement); and 13 were indeterminate (no variation with regard to present/absent because coders agreed 100% that the variable was absent).<sup>c</sup> Information on range, mean, and mode is based on using the final data set with discrepancies between coders resolved via discussion to achieve consensus.

TABLE 3 | EXAMPLE FROM CODING SYSTEM/FORM FOR OPEN-ENDED KNOWLEDGE ITEMS<sup>a</sup>

Knowledge Assessment Question: Describe the Elements of the Model for Improvement	Score <sup>b</sup>
<i>Provides the question and/or one of the following bullets:</i> <ul style="list-style-type: none"> <li>▪ What are we trying to accomplish? (ie, makes reference to an aim, goal, or objective) <ul style="list-style-type: none"> <li>▪ Reference to identification of deficiency is adequate</li> <li>▪ Simply stating identify gap or problem is not sufficient</li> </ul> </li> </ul>	0 or 1
<i>Provides the question and/or one of the following bullets:</i> <ul style="list-style-type: none"> <li>▪ How will we know that a change is an improvement? <ul style="list-style-type: none"> <li>▪ Makes reference to measurement or data collection</li> </ul> </li> </ul>	0 or 1
<i>Provides the question and/or one of the following bullets:</i> <ul style="list-style-type: none"> <li>▪ What change can we make that will result in improvement? <ul style="list-style-type: none"> <li>▪ Makes reference to an intervention, solution, or plan</li> </ul> </li> </ul>	0 or 1
Identifies/mentions plan-do-study/check-act	0 or 2
<i>In addition: 1 point for each step that is defined (must name or reference each step):</i> <ul style="list-style-type: none"> <li>▪ Plan = establish improvement process with goals</li> <li>▪ Do = implement the plan, including outline of responsibility of team members</li> <li>▪ Study = analyze the results from the implemented plan</li> <li>▪ Act = proceed with the plan, modify the plan, or abort the plan</li> </ul>	0 or 1 0 or 1 0 or 1 0 or 1

<sup>a</sup> Instructions to coder for part B, open-ended knowledge items: In coding and scoring this section, each response is read to determine whether the answer addresses the bullet(s) for each variable.

<sup>b</sup> Each relevant unit of information (variable) is scored as present or absent (dichotomous) with some variables weighted more (eg, a score of 2, rather than 1), given their importance or complexity.

### Development of the Coding System for Qualitative Parts

For transforming the qualitative responses into quantifiable scores, we developed a coding system (1) that was comprehensive regarding content, (2) that could be used for assessing novice to expert, (3) that would be able to provide specific feedback to all curriculum stakeholders, (4) that had nonoverlapping variables for coding (a system where the variables and the categories for coding are as exhaustive and mutually exclusive as possible),<sup>24</sup> and (5) that was flexible enough to accommodate variability in project descriptions.

To develop the coding system, we used input from an expert panel along with curriculum goals and relevant literature to compile ideal and comprehensive responses to the questions. Specifically, one of us (A.M.T.) asked 5 additional nationally recognized experts in QI principles (expert panel of leaders and/or experts on QI teaching and practice, with all having published in the area) to complete and comment on the questionnaire. Their comments were positive and served as additional content and face validity. In addition, 4 individuals who were familiar with PBLI principles (but not experts) were asked to complete the tool to provide additional examples for the evolving coding (scoring) system.

The coding system we created scored each relevant unit of information (variable) as present or absent

(dichotomous) and gave greater weight to some items, given their importance or complexity (see TABLES 2 through 4). TABLE 2 summarizes the number of variables and points for both the application and the knowledge components. TABLES 3 and 4 are sections from the coding forms related to knowledge (elements of the Model for Improvement) and application (background and aim sections).

### Resident Participants

Our sample encompassed 86 residents in postgraduate years (PGYs) 1 through 3 (PGY-1, n = 33; PGY-2, n = 23; PGY-3, n = 30) in the Internal Medicine Residency program at University Hospitals of Cleveland and the Louis Stokes Cleveland Department of Veterans Affairs Medical Center (LSCDVAMC). Each year residents complete a 1-month ambulatory rotation where they participate in a variety of specialty medical clinics. Most of this rotation occurs at the LSCDVAMC. From 2004 to 2006, residents participated in either a PBLI curriculum or a systems-based practice curriculum. The curricula were offered on alternate blocks and required for all physician trainees providing ambulatory care. Limited faculty to teach the PBLI curriculum prevented separating residents by level and offering a curriculum tailored to different levels of experience. All 3 levels were involved in the PBLI sessions to ensure exposure to the PBLI. Before the first session for 11 consecutive

TABLE 4 EXAMPLE FROM CODING SYSTEM/FORM FOR OPEN-ENDED APPLICATION ITEM (PROPOSED PROJECT)<sup>a</sup>

Application Assessment Item (Scoring descriptions for the background and aim sections)	Score <sup>b</sup>
1. Background	
A. Provides a summary statement of problem <ul style="list-style-type: none"> <li>Clearly describes what the problem or issue is that is in need of change or provides a specific focus for more exploratory examination. A reference or statement about a solution is not sufficient.</li> </ul>	0 or 1
B. Mentions why this problem (project) is important. <i>Discusses why the problem exists, the current situation, and what is causing the problem to happen.</i> <ul style="list-style-type: none"> <li>Describes the current <i>specific</i> situation (not global statements) and demonstrates understanding of why it is a problem, not just a restatement of the problem or solution.</li> </ul>	0 or 1
C. Mentions consequences (clinical impact) of the current state of affairs or status quo. <ul style="list-style-type: none"> <li>Appropriate response assists in completing the following statement: The presence of this problem causes or leads to the following clinical ramifications. It taps into health consequences, not simply the patient or clinician doing something different. Also, a general reference to catchphrases, such as “falling through the cracks,” is not specific enough, eg, problems related to missed medication should state patient clinical impact when medication is missed, such as hyperglycemia resulting in patient confusion, rather than simply stating that missing medications is bad.</li> </ul>	0 or 2
D. Suggests that more analysis of problem/situation needed before proceeding <ul style="list-style-type: none"> <li><i>Description of measure:</i> Measure relates to identifying the extent of a problem or situation that may require further evaluation or improvement. The description answers the question: How will you know if the problem is significant and needs to be modified/improved? ie, not just predata, but data that informs development of an intervention (it is likely if pre-post design and no language about informing the intervention; it is relevant to evaluating the intervention not assessing the problem).</li> <li>Measures may include qualitative (eg, following a patient’s steps through the process) and quality improvement methodology, such as interviews, flowcharts, fishbone diagrams, Pareto charts, etc.</li> </ul>	0 or 1
2. Aim	
A. Addresses when the improvement will occur (when is it changing) <ul style="list-style-type: none"> <li>Includes specific timeframe for improvement (eg, anticipated gains will be seen within 6 mo).</li> </ul>	0 or 2
B. Addresses how much improvement will result (how much is it changing) <ul style="list-style-type: none"> <li>Identifies what amount of change (quantitative or qualitative) is expected (eg, “50% improvement”)</li> </ul>	0 or 2

<sup>a</sup> Instructions to coder for part A, proposed project: In coding and scoring this section, there are 6 domains that are expected to be addressed: background, aim, measurement, intervention, impact, and next steps. In addition to appropriately addressing the key aspects of these domains, the response in each domain should relate to each other (eg, appropriate measure given aim). The responses are evaluated for whether the bullets are addressed to receive credit for that item (variable).

<sup>b</sup> Scoring of each relevant unit of information (variable) is scored as present or absent (dichotomous) with some variables weighted more (eg, a score of 2 rather than 1), given their importance or complexity.

blocks starting in July 2005, residents were asked to complete the assessment tool. The study was approved and deemed exempt by the LSCDVAMC Institutional Review Board.

### Training for Coders

Training and coding were done in an iterative fashion because the system was evolving. Questionnaires completed by the 86 residents were randomly divided into 7 sets. The 2 coders (the authors), blinded to the identity of the residents, independently scored the knowledge and application components of the questionnaire in a set, then discussed coding decisions to address all discrepancies before continuing to the next set. This iterative process permitted evaluating reliability and also helped to further

refine examples for the coding system. Consensus was achieved for all discrepancies before doing any analyses on total scores.

### Data Analysis Strategy

**Interrater Reliability** Given the iterative nature of developing the coding system, results for the final 2 sets are presented (n = 21). Interrater agreement was evaluated using the exact percentage of agreement and the Cohen  $\kappa$ .<sup>24,25</sup> The  $\kappa$  corrects for chance agreement but is problematic with extreme distributions (eg, for variables that are not observed frequently, such as for samples with no or little training).<sup>24</sup>

Interrater covariation for total knowledge and for total application scores was evaluated using the Lin concordance

correlation (unlike the Pearson  $R$ , this measure takes systematic coding errors into account).<sup>24</sup>

**Validity** In addition to the evaluation of content validity via expert review, we evaluated known-groups validity (discriminative validity) and validity of the underlying ordering of levels of difficulty.

For discriminative validity, we compared 2 groups (those with no prior PBLI experience versus those with prior experience) on their total application score, their scores on self-efficacy items, and their total knowledge score using independent sample  $t$  tests ( $P < .05$ ).

For evaluation of the developmental progression, a scalogram analysis (coefficient of reproducibility) was conducted, which evaluates ordered, internal consistency (ie, pattern of the variables based on difficulty level).<sup>26</sup> This is in contrast to a homogeneity model, where items are developed to be highly correlated with each other (eg, internal consistency as measured by Cronbach  $\alpha$ ). Coefficients of reproducibility were calculated for the 3 domains: (1) the 6 application variables, (2) the 4 self-efficacy items, and (3) the 6 knowledge items. Coefficients need to be 0.90 or higher for evidence of the progression.

## Results

### Reliability

For knowledge and application variables, percentages of agreement were greater than 90%. Correcting for chance agreement identified some variables that are potentially problematic (eg, presence of the variable didn't occur often, and when it did, the coders did not always agree). Although concordances for total scores were high, there is room for improvement in the application component (TABLE 2).

### Validity

Sixty-three residents had no prior experience, and 23 had participated in a prior PBLI curriculum. Although the mean score for residents who had experienced a prior PBLI curriculum was higher for all comparisons, the comparisons regarding teaching CQI and proposed project/application scores were not significant (TABLE 1).

To evaluate the proposed developmental sequence of self-efficacy, ratings of *agree* and *strongly agree* were coded as evidencing efficacy with all other ratings indicating lack of efficacy for each of the 4 items. The patterns were evaluated, and the coefficient of reproducibility was 0.92, indicating strong internal consistency of ordering (items are in the correct order in TABLE 1).

To evaluate the difficulty and sequencing of knowledge items, dichotomous definitions of each of the 6 items were created (no knowledge versus some knowledge). The patterns were evaluated, and the sequence of difficulty (from easiest to most difficult) was some understanding of change concept, Model for Improvement cause-effect diagram, special cause variation, common cause variation,

and why the distinction between the 2 types of variation is important. The coefficient of reproducibility was 0.93, indicating strong consistency in ordering by difficulty.

For the application component (dichotomized no skill versus some application skill evidenced for each of the 6 sections), the sequence of difficulty (from easiest to most difficult) was background, intervention, measures, aim, impact, and next steps. The coefficient of reproducibility was 0.96, indicating strong consistency in ordering by difficulty.

## Discussion

As the field evolves its thinking and expectations about the ACGME residency requirements and potential requirements for practicing physicians, it's worth thinking about assessment methods that anticipate shifts in understanding and experience and that provide direction for nurturing residents to become proficient in PBLI, to engage in CQI, and to become competent teachers of the discipline. To this end, we developed an assessment tool (questionnaire and coding system) consistent with the proposed developmental progression that demonstrated good psychometric properties and has the potential to be useful for formative and summative evaluations.

Although the residents' ranges of prior experiences were small, the tool differentiated between the 2 groups in these preliminary analyses, suggesting we are achieving our desired goals for the tool. The ranges of observed responses for those with prior PBLI experience are low on our scale, developed with ideal responses in mind, but this may reflect reality: prior PBLI experience for these residents was on average a year earlier and had involved a 4-week ambulatory curriculum meeting once a week and working on a team project between weekly meetings. Their experience was minimal, and our results are consistent with what one would expect on a tool designed to differentiate various levels of experience and expertise.

As a tool for formative evaluation,<sup>5</sup> given the infancy of this requirement and the curricula to address it, low scores may be an accurate representation of our developmental level as a field. Our findings suggest that limited curriculum experiences and/or hoping that residents may be acquiring the competency components from general training and experiences may be a comfortable, but unrealistic, belief. Moving toward more systematic training for medical students is consistent with thinking of the full continuum of comprehensive training.<sup>27,28</sup> In addition, such approaches may help address the ACGME's goal of residents being able to contribute to the education of students, other residents, and other health care professionals as part of the PBLI competency. We believe teaching and imparting PBLI knowledge is part of the spirit of ACGME's vision and should begin during residency. If we truly want practicing physicians who are proactive about QI and competent enough to teach PBLI, we need to invest more resources in

PBLI curricula and have detailed, comprehensive assessments.<sup>3,27,29</sup> We need to continue thinking about challenges and solutions to expand and improve curricula, rather than simplifying assessment strategies.

Our study has several limitations. It involved a single residency program at 1 institution, and our sample size and limited levels of experience weren't large enough to do Rasch analyses or more extensive analysis of items.<sup>30</sup> At the same time, our study has one of the larger sample sizes to date, and our initial findings are encouraging.

Another limitation is that only 2 coders were involved in the questionnaire and coding system development. Nevertheless, our questionnaire and coding system were derived with input from national and regional PBLI experts and built on the work of others. Although the tool is successful at creating scoring formats based on ideal responses, the relative inexperience of the sample for this study resulted in a limited range of responses for coding, and thus, it was hard to train on some items (eg, impact and next steps). This was reflected in a few items that had less desirable agreement properties after correcting for chance. Although it is clear additional study is needed, the questionnaire and scoring system perform well overall, especially when considering concordance of total scores. Next steps consist of further evaluating the tool's ability to measure change, to differentiate various and broader levels of experience and settings, and to predict relevant criteria. At this time, it isn't possible to evaluate predictive validity and/or to determine what cut points might be associated with achieving competency. However, as more data become available, including longitudinal databases that include postresidency information, it will be worthwhile to determine scores (overall and by domain) that indicate comprehension, command, and application of the concepts (pass) versus insufficient command. Such cut points will help to target and tailor training to ensure PBLI competency (this is also consistent with Rasch modeling approaches).

The largest potential trade-off to our assessment tool's detailed scoring system may be time for training, which was hard to assess during this developmental phase. After training, we found estimates to range for coding knowledge responses from 1 to 8 minutes and for application responses, from 1 to 7 minutes. Nevertheless, the benefit is our scoring approach provides specific information at the individual trainee level about successes and areas for further improvement particularly when coupled with our previous assessment work of PBLI project impact on clinical operations and trainee satisfaction.<sup>18</sup> Such detailed and comprehensive approaches that inform formative and summative evaluations will be helpful in identifying what ways a curriculum or program is successful and in figuring out reasons why it may not be as effective as anticipated.<sup>31</sup> In an effort to provide more detailed feedback without necessarily providing total scores, our coding form will be

simplified to create a checklist to use as documentation in portfolios to demonstrate increasing ability. This information can also be interpreted within institutions for benchmarking purposes. In addition, our application skills coding tool is appropriate for scoring actual projects and presentations.

Our goal for quantifying the open-ended responses is to be able to provide more specific guidance to residents in their process of obtaining PBLI competency and becoming proactive about QI. Accordingly, although the total score is useful for research and overall resident performance and curriculum evaluation, it is less useful for nurturing development and providing information about specific components or areas of strengths and weaknesses. That is why we attempted to create a more detailed coding system for knowledge and application skills and to avoid global ratings on the part of the coders because global ratings (even several related to different components) may not provide enough information to identify the gaps in knowledge and application and may limit our ability to provide feedback. Although most of our knowledge and application items overlap with others' foci and help to underscore growing convergence about the components of the fundamental skill set for PBLI competency, some approaches use ratings that combine pieces of information into 1 rating that may not be as helpful for those residents whose responses or projects are not scored *excellent* or *outstanding* and may not be as helpful for knowing how to improve the curriculum.<sup>eg, 15–17,32</sup> It is worth noting that these assessment tools, especially for projects, are most likely combined with face-to-face interaction and feedback, which may balance some of these concerns and perhaps our approach may be more helpful for those starting a curriculum.

## Conclusions

Our approach to assessing PBLI application skills, knowledge, and self-efficacy demonstrated good psychometric properties. Although additional evaluation is needed, the SQI TAT may prove helpful, especially in contexts where more detailed information about trainees' progress and the curriculum is desired. Our approach helps to inform discussions as the field continues to tackle the challenge of translating PBLI competency requirements for all physicians into measurable expectations that achieve the goal of empowering physicians to improve health care systems.

## References

- 1 Accreditation Council for Graduate Medical Education. Common program requirements: general competencies. Available at: <http://www.acgme.org/outcome/comp/GeneralCompetenciesStandards21307.pdf>. Approved February 13, 2007; accessed April 17, 2010.
- 2 Accreditation Council for Graduate Medical Education. Facilitator's manual practical implementation of the competencies ACGME, April 2006. [http://www.acgme.org/outcome/e-learn/FacManual\\_module2.pdf](http://www.acgme.org/outcome/e-learn/FacManual_module2.pdf). Accessed April 17, 2010.

- 3 Stevens DB, Sixta CS, Wagner E, Bowen JL. The evidence is at hand for improving care in settings where residents train. *J Gen Intern Med*. 2008;23(7):1116–1117.
- 4 Lynch DC, Swing SR, Horowitz SD, Holt K, Messer JV. Assessing practice-based learning and improvement. *Teach Learn Med*. 2004;16(1):85–92.
- 5 Swick S, Hall S, Beresin E. Assessing the ACGME competencies in psychiatry training programs. *Acad Psychiatry*. 2006;30(4):330–331.
- 6 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.
- 7 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.
- 8 Windish DM, Reed DA, Boonyasai RT, Chakraborti C, Bass EB. Methodological rigor of quality improvement curricula for physician trainees: a systematic review and recommendations for change. *Acad Med*. 2009;84(12):1677–1692.
- 9 Holmboe ES, Prince L, Green M. Teaching and improving quality of care in a primary care internal medicine residency clinic. *Acad Med*. 2005;80(6):571–577.
- 10 Hildebrand C, Trowbridge E, Roach MA, Sullivan AG, Broman AT, Vogelman B. Resident self-assessment and self reflection: University of Wisconsin-Madison's five-year study. *J Gen Intern Med*. 2009;24(3):361–365.
- 11 Coleman MT, Nasraty S, Ostapchuk M, Wheeler S, Looney S, Rhodes S. Introducing practice-based learning and improvement ACGME core competencies into a family medicine residency curriculum. *Jt Comm J Qual Saf*. 2003;29(5):238–247.
- 12 Shunk R, Dulay M, Julian K, et al. Using the American board of internal medicine practice improvement modules to teach internal medicine residents practice improvement. *J Grad Med Educ*. 2010;2(1):90–95.
- 13 Weingart SN, Tess A, Driver J, Aronson MD, Sands K. Creating a quality improvement elective for medical house officers. *J Gen Intern Med*. 2004;19:861–867.
- 14 Fussell JJ, Farrar HC, Blaszk RT, Sisterhen LL. Incorporating the ACGME educational competencies into morbidity and mortality review conferences. *Teach Learn Med*. 2009;21(3):233–239.
- 15 Djuricich AM, Ciccarelli M, Swigonski NL. A continuous quality improvement curriculum for residents: addressing core competency, improving systems. *Acad Med*. 2004;79(10 suppl):S65–S67.
- 16 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.
- 17 Canal DF, Torbeck L, Djuricich AM. Practice-based learning and improvement: a curriculum in continuous quality improvement for surgery residents. *Arch Surg*. 2007;142(5):479–483.
- 18 Tomolo AM, Lawrence RH, Aron DC. A case study of translating ACGME practice-based learning and improvement requirements into reality: systems quality improvement projects as the key component to a comprehensive curriculum. *Qual Saf Health Care*. 2009;18(3):217–224.
- 19 Peters AS, Kimura J, Ladden JD, March E, Moore GT. A Self-instructional model to teach systems-based practice and practice-based learning and improvement. *J Gen Intern Med*. 2008;23(7):931–936.
- 20 Oyler J, Vinci L, Arora V, Johnson J. Teaching internal medicine residents quality improvement techniques using the ABIM's practice improvement modules. *J Gen Intern Med*. 2008;23(7):927–930.
- 21 Morrison LJ, Headrick LA, Ogrinc G, Foster T. The quality improvement knowledge application tool: an instrument to assess knowledge application in practice-based learning and improvement [abstract]. *J Gen Intern Med*. 2003;18(suppl 1):250.
- 22 Morrison LJ, Headrick LA. Teaching residents about practice-based learning and improvement. *Jt Comm J Qual Patient Saf*. 2008;34(8):453–459.
- 23 Leenstra JL, Beckman TJ, Reed DA, et al. Validation of a method for assessing resident physicians' quality improvement proposals. *J Gen Intern Med*. 2007;22(9):1330–1334.
- 24 Neuendorf KA. *The Content Analysis Guidebook*. Thousand Oaks, CA: Sage Publications; 2002.
- 25 Boyatzis RE. *Transforming Qualitative Information: Thematic Analysis and Code Development*. Thousand Oaks, CA: Sage Publications; 1998.
- 26 McIver JP, Carmines EG. *Unidimensional Scaling*. Newbury Park, CA: Sage Publications; 1981.
- 27 Ogrinc G, West A, Eliassen MS, Liuw S, Schiffman J, Cochran N. Integrating practice-based learning and improvement into medical student learning: evaluating complex curricular innovations. *Teach Learn Med*. 2007;19(3):221–229.
- 28 Huntington JT, Dycus P, Hix C, et al. A standardized curriculum to introduce novice health professional students to practice-based learning and improvement: a multi-institutional pilot study. *Q Manag Health Care*. 2009;18(3):174–181.
- 29 Batalden P, Davidoff F. Teaching quality improvement: the devil is in the details. *JAMA*. 2007;298(9):1059–1061.
- 30 Andrich D. *Rasch Models for Measurement*. Newbury Park, CA: Sage Publications; 1988.
- 31 Ogrinc G, Batalden P. Realist evaluation as a framework for the assessment of teaching about the improvement of care. *J Nurs Educ*. 2009;48(12):661–667.
- 32 Djuricich AM. A continuous quality improvement curriculum for residents. <http://www.aamc.org/mededportal> (ID = 468). Accessed July 21, 2010.

#### APPENDIX SYSTEMS QUALITY IMPROVEMENT TRAINING AND ASSESSMENT TOOL (SQI TAT)

*There are 2 main parts to the tool (part A and part B) presented below without response formatting included (ie, blank spaces for open-ended items and individual rating scales for the self-efficacy items).*

PLEASE COMPLETE THE FOLLOWING QUESTIONS IN PART A AND PART B (PART B IS A SEPARATE PART AFTER COMPLETING PART A AND RETURNING IT). THANK YOU.

#### Part A

Based on your clinical experiences, develop a project that would help to improve any aspects of patient care. Please provide enough information so that someone unfamiliar with the context would know what to do, how to do it, and why.

#### Part B

- I believe I am able to develop a CQI project.\*  
 I believe I am able to implement a CQI project.\*  
 I believe I am able to teach CQI principles.\*  
 I believe I am comfortable developing a data collection plan consistent with time and resource limitations.\*

\* Consistent with the work by Djuricich and colleagues,<sup>15,17</sup> the following response format is used for the above 4 self-efficacy items: strongly agree, agree, don't know, disagree, strongly disagree.

*Please provide a brief response to each of the following:*  
 Describe what is meant by a change concept.  
 Describe how a cause-effect diagram is created.  
 Describe the elements of the Model for Improvement.  
 Define common cause (random) variation.  
 Define special cause (assignable) variation.  
 Why is the distinction between random and special cause variations important?