

# A Scoring Rubric for the Knowledge Section of the Systems Quality Improvement Training and Assessment Tool

Corrine Abraham, DNP, RN\*, Krysta Johnson-Martinez, MD, Anne Tomolo, MD, MPH

\*Corresponding Author: [cabraha@emory.edu](mailto:cabraha@emory.edu)

## Abstract

**Introduction:** Quality improvement (QI) competencies for health professions trainees were developed to address health care quality. Strategies to integrate QI into curricula exist, but methods for assessing interdisciplinary learners' competency are less developed. We refined the Knowledge section scoring rubric of the Systems Quality Improvement Training and Assessment Tool (SQI TAT) and examined its validity evidence. **Methods:** In 2017, the SQI TAT Knowledge section was expanded to cover seven core QI concepts, and the scoring rubric was refined. Three coders independently scored 35 SQI TAT Knowledge sections (18 pretests, 17 posttests). Interrater reliability was assessed by percent agreement and Cohen's kappa for individual variables and by Lin's concordance correlation for total scores for knowledge and application. Concurrent validity was assessed by comparing responses from two groups with different QI exposure and evaluating whether differences in exposure were measured. **Results:** Total-score interrater reliability average measures of concordance were .89 for all coders and  $>.70$  for six of seven concept scores. The total score discriminated the two groups ( $p < .05$ ), and five of seven concept scores were higher for the group with more QI experience. Total scores were significantly higher posttest than pretest ( $p < .001$ ), with improvement in posttest knowledge scores. **Discussion:** The SQI TAT Knowledge section provides a comprehensive assessment of QI knowledge. The scoring rubric was able to discriminate QI knowledge along a continuum. The SQI TAT Knowledge section is not linked to a clinical context, making it useful for assessing interprofessional learners and varying education levels.

## Keywords

Assessment, Competency-Based Medical Education (Competencies, Milestones, EPAs), Curriculum Development, Interprofessional Education, Program Evaluation, Quality Improvement/Patient Safety

## Educational Objectives

By using this assessment tool and scoring rubric, educators will be able to:

1. Objectively assess learner quality improvement knowledge.
2. Assess learner acquisition of quality improvement knowledge along a continuum from simple to complex.
3. Identify learner knowledge gaps and implications for curriculum design (i.e., potential targets for adjusting quality improvement curricula and learning experiences).

## Introduction

Competency in quality improvement (QI) is critical to address the quality and safety gap in health care. Educational programs, accrediting agencies, and nonprofit organizations have published recommendations and requirements to address this need.<sup>1-4</sup> Core competencies and strategies to integrate QI into curricula exist, but generic methods for assessing interdisciplinary learners' achievement of competency are less developed.<sup>5-8</sup> Assessment tools vary in their inclusion of QI principles, depth of the knowledge evaluated, or linkage to clinical experience, making it difficult to identify knowledge gaps for disciplines lacking clinical context.<sup>5-14</sup>

The Quality Improvement Knowledge Application Tool Revised is a widely used scenario-based assessment tool focusing on select foundational QI principles.<sup>8</sup> Its scoring rubric has been revised, which improved interrater reliability; however, limitations remain. For example, the tool fails to assess the breadth and depth of specific QI principles as it was designed for use with novice

### Citation:

Abraham C, Johnson-Martinez K, Tomolo A. A scoring rubric for the knowledge section of the systems quality improvement training and assessment tool. *MedEdPORTAL*. 2022;18:11290. [https://doi.org/10.15766/mep\\_2374-8265.11290](https://doi.org/10.15766/mep_2374-8265.11290)

learners.<sup>8</sup> Scenario-based assessments present challenges for interprofessional learners, and scoring may be influenced by the evaluator’s understanding of the clinical context.<sup>8</sup>

The Systems Quality Improvement Training and Assessment Tool (SQI TAT) assesses QI knowledge, application skills, and attitudes independent of clinical practice.<sup>6</sup> Its Knowledge section includes six open-ended questions with a weighted coding system (rubric) to score each relevant unit of information (variable), with some variables weighted more given the complexity of the concept. The open-ended nature of the Knowledge section focuses on distinct core QI concepts, and the detailed scoring rubric assesses scope and depth of knowledge for each concept.

In 2017, we implemented a Center of Excellence, an innovative model for interprofessional collaborative practice and QI education incorporating a curriculum presented to trainees with varying QI exposure before or during their rotation (i.e., previous educational/QI experiences, rotation length, and involvement in center projects). The core curriculum included five sessions based on the Model for Improvement that provided instruction on the foundations of quality, including common methods and skills and case-based application (Table 1).<sup>15</sup> The curriculum was not explicitly linked to components of the SQI TAT knowledge assessment; therefore, some concepts were not fully addressed (statistical process control in particular). Some trainees had the opportunity to apply their knowledge to a QI project within the center, receiving one-on-one coaching. The

projects were at differing phases of completion (i.e., problem analysis, implementation, or evaluation) depending upon the time frame for the clinical rotation.

We sought to understand the development of QI knowledge among the different trainees and to evaluate any relationship to QI exposure and potential gaps in the curriculum. The validity evidence for use of the original scoring rubric to assess interprofessional trainees participating in a clinical rotation had never been evaluated. This highlighted the need to refine the SQI TAT knowledge scoring rubric to differentiate knowledge acquisition over time and exposure to curriculum or experiences. Here, we describe the refinement of the SQI TAT scoring rubric criteria and examination of the rubric’s validity evidence, providing a tool for QI learner assessment.

### Methods

In 2017, two QI experts were recruited to assist the SQI TAT developers with revising the Knowledge section and scoring rubric (Appendices A and B). Each was a health professional and clinician with advanced training and extensive experience in QI; in curriculum design for staff, trainees, and faculty; and in teaching trainees with variable levels of clinical and QI experience. They had completed the same 2-year national postdoctoral fellowship focused on developing QI competencies where they led system-level QI projects. Both experts had also completed specialized training in Lean and Six Sigma methods.<sup>15</sup> After finishing the fellowship, they remained actively engaged as mentors and coaches, assisting to establish a local QI curriculum

**Table 1.** QI Curriculum: Specialty Care Education Center of Excellence<sup>a</sup>

Session	Session Objectives	Tools
1	<ol style="list-style-type: none"> <li>1. Describe personal and program objectives for QI learning and development.</li> <li>2. Discuss the case for quality as related to practice area of focus/interest.</li> <li>3. Outline improvement framework—purpose and analysis of problem.</li> <li>4. Assess clinical microsystem using the clinical microsystem 5 P’s guide.</li> </ol>	<ul style="list-style-type: none"> <li>• Clinical microsystem assessment</li> <li>• Process map</li> <li>• Cause-and-effect diagram</li> </ul>
2	<ol style="list-style-type: none"> <li>1. Analyze data to determine an area of focus for a QI project.</li> <li>2. Identify types of measures (outcome, process, and balancing measures) using the clinical value compass.</li> <li>3. Develop conceptual and operational definitions for a QI case.</li> <li>4. Outline a data collection plan.</li> </ol>	<ul style="list-style-type: none"> <li>• Pareto chart</li> <li>• Clinical value compass</li> </ul>
3	<ol style="list-style-type: none"> <li>1. Apply QI methods: <ul style="list-style-type: none"> <li>• Based upon analysis of problem, create a SMART aim statement.</li> <li>• Identify change concepts pertinent to analysis of process.</li> </ul> </li> <li>2. Human factors engineering and reliability science: <ul style="list-style-type: none"> <li>• Identify vulnerabilities to human performance.</li> <li>• Apply human factors engineering principles in developing interventions (strong, medium, and weak interventions).</li> </ul> </li> </ol>	<ul style="list-style-type: none"> <li>• SMART aim</li> <li>• Change concept</li> <li>• Human factors engineering</li> <li>• Action hierarchy</li> <li>• Data collection plan</li> </ul>
4	<ol style="list-style-type: none"> <li>1. Apply the Model for Improvement to your QI project.</li> <li>2. Describe components of the PDSA cycle.</li> <li>3. Identify implications for sustaining and spreading an intervention to a clinical case.</li> </ol>	<ul style="list-style-type: none"> <li>• PDSA cycle</li> <li>• Implementation plan</li> </ul>
5	<ol style="list-style-type: none"> <li>1. Identify components of a run chart.</li> <li>2. Determine types of variation based on rules for detecting special cause signals.</li> <li>3. Interpret findings to identify next steps.</li> </ol>	<ul style="list-style-type: none"> <li>• Median, range, runs, astronomical points</li> <li>• Common and special cause variation</li> <li>• Adopt, adapt, abandon</li> </ul>

Abbreviations: 5 P’s, purpose, patients, professionals, processes, patterns; PDSA, plan/do/study/act; QI, quality improvement; SMART, specific, measurable, achievable, relevant, time. <sup>a</sup>Implementation and application within a project varied by duration of clinical rotation (4-12 weeks).

for interprofessional learners as well as collaborating to develop and implement a faculty development program. After completing the fellowship, both secured system-level leadership positions, remained active in quality initiatives in their organizations, and led development and implementation of QI curricula for varied audiences.

#### Training the Raters Through Iterative Review

Over the course of approximately seven sessions, the developers trained the new coders by reviewing interprofessional trainee responses and then convening to discuss areas with scoring variability and gaps in the scoring tool. The group rated previously completed Knowledge section responses from a convenience sample of interprofessional trainees who had participated in disparate QI experiences (2016-2017). The two new coders and one of the developers (Corrine Abraham, Krysta Johnson-Martinez, and Anne Tomolo) were masked to respondent identity and independently scored responses using the scoring criteria and examples included on the rubric. All decisions were dichotomous—either the response met criteria or not. The remaining developer facilitated a discussion of each item, our criteria for scoring, and our rationale for scoring decisions. We reconciled discrepancies and collaborated to refine the wording, descriptions, and clarity of the rubric, incorporating specific example responses. We repeated this process in an iterative fashion, rating a portion at a time independently, discussing, and reconciling differences.

To ensure comprehensive content, we decided to separate one multicomponent question to prompt more targeted responses (indicated with an \* in Appendix A). To decrease variability in scoring, we added details to the rubric to assess knowledge of the SMART acronym (specific, measurable, achievable, relevant, time) more completely in the aim statement. The team added specific criteria with examples demonstrating knowledge acquisition from basic to more complex to the rubric to enhance scoring clarity.<sup>16</sup> The team weighted some variables more to reflect concepts of increased complexity (Table 2). The final scoring rubric includes seven QI concept scores and a total possible score of 49 resulting from 33 dichotomous variables. Points scored on the rubric reflect the presence of knowledge (yes/no) and depth whereby more points equal greater knowledge. The rubric was designed to reflect knowledge attainment along a continuum, with interpretation of scores dependent upon the learning outcomes desired. Iterative use of the rubric to score new responses served to train the new coders, guided further rubric revisions, and permitted evaluation of interrater reliability.

#### Evaluating the Scoring Rubric's Validity Evidence

To evaluate the validity evidence of the final version of the scoring rubric, we independently scored 35 new SQI TAT Knowledge sections (18 pretests and 17 posttests) from 18 learners. All learners participated in a specialty clinical rotation in the Center of Excellence that incorporated a QI curriculum varying in duration and depth of content and application. Before and after participating, all learners received an online link to the SQI TAT with instructions to complete the questions to the best of their ability without the use of external resources. We reiterated that the results would remain confidential, would not impact learners' course evaluation, and would be used to help us improve the educational experience. The PRAM (Program for Reliability Assessment with Multiple Coders) package was used to conduct all reliability analyses.<sup>17</sup> SPSS Statistics version 24.0 was used for the validity and change analyses.

*Interrater reliability:* We evaluated interrater agreement using average of exact percent agreement (all possible comparisons are averaged) and Cohen's kappa for multiple coders (Appendix C). Kappa corrects for chance agreement but is problematic for infrequently occurring variables.<sup>17</sup>

The team resolved discrepancies using the score agreed upon by two of the three coders. We evaluated interrater covariation for total knowledge and for each QI concept using Lin's concordance correlation based on the expert consensus as proxy for the gold standard. Lin's measure also takes systematic coding errors into account and is robust with smaller sample sizes.<sup>17</sup>

*Relation to other variables:* The team assessed concurrent validity by comparing two groups to determine if the scoring rubric differentiated the groups consistent with the expectation that higher scores would be related to more experience. The two groups were (1) those with no or limited prior QI experience (had only attended a meeting discussing QI) versus (2) those with prior experience (had attended a teaching session) or who had been a passive or active part of a QI team. We compared the groups on their total knowledge score and seven QI concept scores at baseline using independent samples *t* tests ( $p < .05$ ).

*Measuring change:* The team assessed the ability of the scoring rubric to measure change over time in individual respondent scores for trainees having completed both a pretest and posttest ( $n = 15$ ). We compared total knowledge and seven QI concept scores using the paired *t* test ( $p < .05$ ), with the expectation that posttest scores would be higher than pretest scores.

**Table 2.** SQI TAT Knowledge Assessment Scoring Criteria

Concept/Question	Variables	Scoring/Weight	Concept Total Score		
Aim statement: What are the key elements of an aim statement?	SMART acronym	1	7		
	Specific	1			
	Measurable	1			
	Achievable	1			
	Relevant	1			
	Relevant—stakeholder or organization mission	1			
	Timely	1			
Change concept: Describe what is meant by a change concept, and give an example related to health care.	Define—developing	1	6		
	Define—creative thinking	1			
	Example theme	1			
	Developed example	3			
Cause-and-effect diagram: Describe how a cause-and-effect diagram is created.	Describes process	1	6		
	Basic definition	1			
	Main effect and primary causes	2			
	Describes subfactors	2			
Model for Improvement: Describe the elements of the improvement model.	What are we trying to accomplish?	1	8		
	How will we know a change is an improvement?	1			
	What changes can we make?	1			
	PDSA—list (or spell out words)	1			
	Define plan	1			
	Define do	1			
	Define study	1			
	Define act	1			
	Common cause variation: Define common cause (random) variation.	Natural part of process		2	5
		Sum of small variations		2	
No single root cause		1			
Easily identified problem		2			
Special cause variation: Define special cause (assignable) variation.	Not inherent—unexpected	2	5		
	Single root cause	1			
	Distinction between common cause and special cause variation: Why is the distinction between random and special cause variation important?	Address special cause first		3	12
Special cause detected, then eliminate or adopt		3			
After special cause, then common cause addressed		3			
Common cause reduced by changes to process		3			
Total SQI TAT Knowledge assessment score			49		

Abbreviations: PDSA, plan/do/study/act; SMART, specific, measurable, achievable, relevant, time; SQI TAT, Systems Quality Improvement Training and Assessment Tool.

**Ethical Considerations**

The SQI TAT was introduced as part of the QI curriculum for trainees to evaluate their own learning and growth in QI knowledge and skills. We did not share individual results with training program faculty; therefore, scores were not used to evaluate performance. QI faculty serving as coders deidentified the SQI TAT responses to evaluate the curriculum and guide refinement of the tool.

**Results**

**Trainee Characteristics**

The interprofessional cohort who completed the SQI TAT included social work interns (*n* = 3), nurse practitioner residents (*n* = 4), physician assistant residents (*n* = 3), palliative care fellows (*n* = 1), preventive medicine residents (*n* = 2), psychiatry residents (*n* = 2), pre-MPH practicum learners (*n* = 1), and postdoctoral fellows (*n* = 2). At baseline, experience was weighted toward those with no previous experience (*n* = 8). Of

those with experience, two had attended a meeting where QI was discussed, three had attended a lecture or teaching session, and five were an active part of a QI team.

**Validity Evidence**

*Interrater reliability for individual variables:* Average percent agreement (all coders agreed) was above 80% for 32 of the 33 variables. The exception was an average percent agreement of 73% for one change concept variable. The results for Cohen’s kappa for multiple coders included the following: Thirteen variables were .75 or higher (indicating excellent agreement beyond chance), 14 were between .40 and .74 (indicating fair to good agreement beyond chance), four were below .40 (poor agreement), and two were indeterminate (no variation regarding presence/absence because all coders agreed that the variable was absent; Appendix C).

*Total and QI concept scores:* Table 3 summarizes the results for average Lin’s concordance correlation values. All but two

**Table 3.** Summary of Interrater Reliability for Concepts and Total Score Concordance

Knowledge Concept	No. of Variables Coded	No. of Points Possible	Average Lin's Concordance Value	Highest Concordance Value (Pair)	Lowest Concordance Value (Pair)	Difference in Concordance Values
Elements of an aim statement	7	7	.91	.93 (2, 3)	.89 (2, 1)	.04
Meaning of change concept	4	6	.73	.76 (2, 1)	.68 (2, 3)	.08
Description of creation of cause-and-effect diagram	4	6	.89	.95 (2, 3)	.88 (3, 1)	.07
Description of elements of the improvement model	8	8	.91	.94 (2, 3)	.88 (2, 1)	.06
Meaning of common cause (random) variation	3	5	.70	.75 (3, 1)	.68 (2, 3)	.07
Meaning of special cause (assignable) variation	3	5	.69	.72 (2, 1)	.69 (2, 3)	.03
Understanding of distinction between the two variations	4	12	.50	.64 (3, 1)	.39 (2, 3)	.25
Total	33	49	.89	.91 (3, 1)	.86 (2, 1)	.05

were .70 or higher, with one of those two being .69. Total-score concordance correlation was .89. The difference between the highest and lowest Lin's concordance values for each concept demonstrates the level of variation between coders compared two at a time. The lowest concordance value (.50) was for understanding the distinction between the two types of variation. This item also had the largest difference (.25) between high and low concordance correlations (for all the other scales, the difference was between .03 and .08).

*Relation to other variables:* Ten trainees had no prior experience or had only attended a meeting, and eight had participated in a prior QI session or were involved in a project. The mean pretest score for trainees who had participated in a QI course was higher for seven of the eight comparisons (Table 4). Total score discriminated the two groups ( $p < .05$ ), and the pattern of scores for five of the seven QI concept scores was consistent with level of experience. Meaning of change concept and the distinction between two types of variation were inconsistent, indicating low

knowledge levels regardless of the group. Of the total knowledge score possible (49), the group with more experience had a mean of 9.7, with a range from 21 to 0 (one learner scored 0). The group with no experience had a mean of 4.5, with a range from 15 to 0 (four learners scored 0).

*Measuring change:* Posttest total scores were significantly higher than pretest total scores ( $p < .001$ ). The pattern of scores for all seven QI concept scores was consistent with improvement in posttest knowledge (Table 3), with three concepts being statistically significant (aim statement, cause-and-effect diagram, and improvement model knowledge). Knowledge about change concepts and variation did not significantly change.

### Discussion

To assess the development of QI knowledge among interprofessional trainees and to evaluate potential gaps in the curriculum, we refined the SQI TAT scoring rubric criteria and examined the rubric's validity. Specific examples were refined or added for each variable to enhance consistent scoring of the dichotomous variables. Agreement among coders with a shared mental model yielded consistency on 32 of 33 variables of the SQI TAT scoring rubric. The expansion of the SMART aim concept provided a more complete assessment of basic QI knowledge. Through an iterative process, we were able to establish and refine meaningful validity evidence for our scoring rubric for assessing fundamental QI knowledge.

The SQI TAT scoring rubric can be used for assessing novice to more advanced learners, has discrete criteria for scoring, and can accommodate variability in responses. The rubric includes criteria to decrease the subjective nature of evaluating open-ended responses and weighted scoring to detect level of

**Table 4.** Comparison of Baseline Knowledge Scores for Residents Who Had No or Limited Prior Experience ( $n = 10$ ) Versus Residents Who Had Prior Experience ( $n = 8$ )

Scores	No Prior Experience: <i>M (SD)</i>	Prior Experience: <i>M (SD)</i>
Total Knowledge	4.5 (5.7)	9.7 (7.5) <sup>a</sup>
Subscales		
Elements of an aim statement	0.6 (0.8)	2.4 (2.2)
Meaning of change concept	1.4 (2.3)	1.0 (1.4)
Description of cause-and-effect diagram	0.5 (1.6)	2.4 (2.5)
Improvement model	0.8 (1.7)	1.3 (1.7)
Common cause (random) variation	0.6 (1.0)	1.0 (1.0)
Special cause (assignable) variation	0.6 (1.0)	1.6 (1.8)
Importance of distinction between the two variations	0.0 (0.0)	0.0 (0.0)

<sup>a</sup>Difference significant at  $p < .05$ . Information is based on using a data set with discrepancies resolved by utilizing the value agreed on by two of the coders.

expertise. The SQI TAT Knowledge section is not based upon self-assessment or clinical context, making the tool objective and adaptable for learners from varied professions and educational levels. The tool gauges development of QI knowledge across a learning continuum, which is useful formatively to assess growth and summatively to assess learning outcomes. As a curricular evaluation tool, it can be used to identify challenging concepts for targeting curricular enhancements. Knowledge acquisition is one component of developing competency in QI, and this rubric is a structured way to evaluate attainment of QI knowledge. Attitude and application of QI concepts are evaluated in the other sections of the SQI TAT, providing a well-rounded evaluation of QI competency.<sup>6</sup>

We gained useful insights using the SQI TAT in evaluating our QI curricula. Our results indicated noticeable improvement in novice trainees and a relationship between higher scores and greater QI exposure. Due to the lack of more advanced learning opportunities (curricular and situated learning), we were limited in validating the ability of the scoring rubric to differentiate complex knowledge development, that is, concepts linked to system redesign (change) and analytics (variation). Application of these concepts occurs more intermittently and may depend upon learners' unique experiences in applying the concepts to an actual QI project. For example, projects progress differently, and not all learners have robust data with which to gain an understanding of measurement and analysis. Identification of these gaps in knowledge highlighted areas in the curriculum in need of more reinforcement independent of situated learning opportunities.

Our findings assessing simple to complex knowledge acquisition were limited due to the sample size, variable QI exposure, and timing for completing the survey postcourse. We did not standardize and control administration (e.g., how soon after completion of the course/project to complete the survey or how much time should be allotted to do so). The motivation for completing the survey was variable as there was no incentive for learners to perform well. Though most items are generic, the SQI TAT Knowledge section aligns with the Model for Improvement rather than other QI methodologies.<sup>15</sup> Utilizing this tool for assessing learners participating in programs aligned with other QI frameworks may not be appropriate if educators desire to evaluate distinct knowledge unique to a methodology (e.g., DMAIC [define/measure/analyze/improve/control] used with Lean or Six Sigma<sup>15</sup>).

The interrater reliability was likely influenced by our similar approach to teaching and implementing QI, leading us to

evaluate the responses from a similar perspective. There were variables with less interrater reliability when using Cohen's kappa for multiple coders. When correcting for chance agreement, most of the variables identified as potentially problematic were related to rare event problems. We still recommend keeping the more advanced concepts as a component of the SQI TAT assessment since they are reflective of expert understanding of QI analytic methods. Due to the rare occurrence of more advanced responses, further evaluation of the scoring rubric for these responses is needed with a more diverse sample.

To use the scoring rubric effectively, raters need to be knowledgeable about basic QI concepts such as those depicted in the Model for Improvement. The rubric provides detailed and specific examples of correct responses for each item as a guide based upon expert consensus. For programs focused exclusively on Lean or Six Sigma methodology, there may be some gaps; however, the rubric includes examples that are generic and fundamental to all QI approaches. Higher scores indicate deeper and broader knowledge of the QI concepts due to the design of the rubric weighting concept variables for complexity. The interpretation of scores is dependent upon the purpose (gauging individual growth or evaluating a curriculum), program outcome objectives, and curricular design.

We anticipate that minimal training is required to correctly score responses; however, we plan to explore how training and expertise of raters may impact the validity of utilizing this tool in different contexts. We plan to further assess the usability and reliability of the scoring rubric controlling for QI experience of respondent, characteristics of QI curriculum, and using external coders with differing QI backgrounds. This will provide more robust validity evidence for the advanced concepts and confirm the extent of training required to use the rubric. Next steps also include developing and testing a scoring rubric for the Application section of the SQI TAT, where learners describe an example of a QI initiative from their unique experience.

The refinement of the SQI TAT Knowledge section provides an assessment of foundational QI knowledge. The scoring rubric can be used for evaluation of individual QI knowledge formatively and summatively, with learners of varying QI exposure, and to detect change in knowledge acquisition over time. It is also a useful tool and rubric for QI educators to strengthen local QI curricula. Attainment of foundational QI knowledge is one component of training health care professionals to improve the environment in which they provide care.



## Appendices

- A. SQI TAT.doc
- B. SQI TAT Part D Scoring Rubric.docx
- C. Scoring Analysis Table.docx

All appendices are peer reviewed as integral parts of the Original Publication.

**Corrine Abraham, DNP, RN:** Associate Professor, Nell Hodgson Woodruff School of Nursing at Emory University; Coordinator, Evidence-Based Practice and Innovation, and Co-Director, VA Quality Scholars Fellowship Program, Atlanta VA Health Care System; ORCID: <https://orcid.org/0000-0002-2160-7894>

**Krysta Johnson-Martinez, MD:** Specialty Care Lead and Chief Medical Informatics Officer, VISN 8 VA Sunshine Healthcare Network

**Anne Tomolo, MD, MPH:** Physician, National Center for Patient Safety; Associate Professor, Emory University School of Medicine

## Acknowledgments

Renee Lawrence, PhD, was instrumental in establishing and implementing the analytic methods to support the development of the Systems Quality Improvement Training and Assessment Tool Knowledge section.

## Disclosures

None to report.

## Funding/Support

This work was supported in part by the Department of Veterans Affairs, Veterans Health Administration, Office of Academic Affiliations Specialty Care Education Center of Excellence grant.

## Prior Presentations

Johnson-Martinez K, Tomolo A, Abraham C. Expanding, refining and evaluating the knowledge component of the Systems Quality Improvement Training and Assessment Tool. Poster presented at: Learn Serve Lead 2019: the AAMC Annual Meeting; November 8-12, 2019; Phoenix, AZ.

## Ethical Approval

Reported as not applicable.

## References

1. Common program requirements. Accreditation Council for Graduate Medical Education. Accessed October 14, 2022. <https://www.acgme.org/what-we-do/accreditation/common-program-requirements/>
2. *Graduate-Level QSEN Competencies: Knowledge, Skills and Attitudes*. American Association of Colleges of Nursing QSEN Education Consortium; 2012. Accessed October 14, 2022. <http://www.aacnnursing.org/Portals/42/AcademicNursing/CurriculumGuidelines/Graduate-QSEN-Competencies.pdf>
3. Association of American Medical Colleges. *Quality Improvement and Patient Safety Competencies Across the Learning Continuum*. Association of American Medical Colleges; 2019. Accessed October 14, 2022. <https://store.aamc.org/quality-improvement-and-patient-safety-competencies-across-the-learning-continuum.html>
4. *The Essentials: Core Competencies for Professional Nursing Education*. American Association of Colleges of Nursing; 2021. Accessed October 14, 2022. <https://www.aacnnursing.org/Portals/42/AcademicNursing/pdf/Essentials-2021.pdf>
5. 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. *J Gen Intern Med*. 2003;18(suppl 1):250. Society of General Internal Medicine Annual Meeting abstract.
6. 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. <https://doi.org/10.4300/JGME-D-10-00102.1>
7. Tomolo AM, Lawrence RH, Watts B, Augustine S, Aron DC, Singh MK. Pilot study evaluating a practice-based learning and improvement curriculum focusing on the development of system-level quality improvement skills. *J Grad Med Educ*. 2011;3(1):49-58. <https://doi.org/10.4300/JGME-D-10-00104.1>
8. Singh MK, Ogrinc G, Cox KR, et al. The Quality Improvement Knowledge Application Tool Revised (QIKAT-R). *Acad Med*. 2014;89(10):1386-1391. <https://doi.org/10.1097/ACM.0000000000000456>
9. Trent P, Dolansky MA, DeBrew JK, Petty GM. RN-to-BSN students' quality improvement knowledge, skills, confidence, and systems thinking. *J Nurs Educ*. 2017;56(12):737-740. <https://doi.org/10.3928/01484834-20171120-06>
10. Doupnik SK, Ziniel SI, Glissmeyer EW, Moses JM. Validity and reliability of a tool to assess quality improvement knowledge and skills in pediatrics residents. *J Grad Med Educ*. 2017;9(1):79-84. <https://doi.org/10.4300/JGME-D-15-00799.1>
11. Brown A, Nidumolu A, McConnell M, Hecker K, Grierson L. Development and psychometric evaluation of an instrument to measure knowledge, skills, and attitudes towards quality improvement in health professions education: the Beliefs, Attitudes, Skills, and Confidence in Quality Improvement (BASiC-QI) Scale. *Perspect Med Educ*. 2019;8(3):167-176. <https://doi.org/10.1007/s40037-019-0511-8>
12. Clarke MJ, Steffens FL, Mallory GW, et al. Incorporating quality improvement into resident education: structured curriculum,

- evaluation, and quality improvement projects. *World Neurosurg.* 2019;126:e1112-e1120.  
<https://doi.org/10.1016/j.wneu.2019.02.214>
13. Steele EM, Butcher R, Carluzzo KL, Watts BV. Development of a tool to assess trainees' ability to design and conduct quality improvement projects. *Am J Med Qual.* 2020;35(2):125-132.  
<https://doi.org/10.1177/1062860619853880>
14. Goodman CW, Justo J, Merrow C, Prest P, Ramsey E, Ray D. An experiential learning collaborative on quality improvement for interprofessional learners. *J Interprof Care.* 2022;36(2):327-330.  
<https://doi.org/10.1080/13561820.2021.1901673>
15. Langley GL, Moen RD, Nolan KM, Nolan TW, Norman CL, Provost LP. *The Improvement Guide: A Practical Approach to Enhancing Organizational Performance.* 2nd ed. Jossey-Bass; 2009.
16. Krathwohl DR. A revision of Bloom's taxonomy: an overview. *Theory Pract.* 2002;41(4):212-218.  
[https://doi.org/10.1207/s15430421tip4104\\_2](https://doi.org/10.1207/s15430421tip4104_2)
17. Neuendorf KA. *The Content Analysis Guidebook.* 2nd ed. SAGE; 2017.

**Received:** March 21, 2022

**Accepted:** September 27, 2022

**Published:** December 13, 2022