Development of a Multi-Domain Assessment Tool for Quality Improvement Projects

Glenn Rosenbluth, MD
Natalie J. Burman, DO, MEd
Sumant R. Ranji, MD
Christy K. Boscardin, PhD

ABSTRACT

Background Improving the quality of health care and education has become a mandate at all levels within the medical profession. While several published quality improvement (QI) assessment tools exist, all have limitations in addressing the range of QI projects undertaken by learners in undergraduate medical education, graduate medical education, and continuing medical education.

Objective We developed and validated a tool to assess QI projects with learner engagement across the educational continuum.

Methods After reviewing existing tools, we interviewed local faculty who taught QI to understand how learners were engaged and what these faculty wanted in an ideal assessment tool. We then developed a list of competencies associated with QI, established items linked to these competencies, revised the items using an iterative process, and collected validity evidence for the tool.

Results The resulting Multi-Domain Assessment of Quality Improvement Projects (MAQIP) rating tool contains 9 items, with criteria that may be completely fulfilled, partially fulfilled, or not fulfilled. Interrater reliability was 0.77. Untrained local faculty were able to use the tool with minimal guidance.

Conclusions The MAQIP is a 9-item, user-friendly tool that can be used to assess QI projects at various stages and to provide formative and summative feedback to learners at all levels.

Introduction

The Accreditation Council for Graduate Medical Education (ACGME) and the Association of American Medical Colleges emphasize quality improvement (QI) and patient safety training in medical education.[1–3] Early data from ACGME Clinical Learning Environment Review (CLER) site visits suggest that academic medical centers struggle to provide "experiential training in all phases of QI"[4] for their residents and fellows.

To accurately document competency, active engagement, and opportunity for feedback, educators need high-quality assessment tools for QI projects. A number of tools exist to assess elements of QI knowledge and practice, each with important strengths and some limitations. For example, the Quality Improvement Knowledge Application Tool (QIKAT)[5] and its revision, the QIKAT-R,[6] are designed to assess knowledge and application of QI principles to projects, and the Systems Quality Improvement Training and Assessment Tool[7] assesses QI skills, knowledge, and self-efficacy, but these tools do not assess learners' QI projects.

DOI: http://dx.doi.org/10.4300/JGME-D-17-00041.1

The Mayo Evaluation of Reflection on Improvement Tool (MERIT)[8] measures critical reflections on QI opportunities, and the Quality Improvement Project Assessment Tool (QIPAT-7)[9] provides an assessment of QI proposals, with a focus on early stages of project development. However, neither the MERIT nor the QIPAT-7 assess implementation, and they do not measure learners' active participation required by the ACGME[2] and the American Board of Medical Specialties Maintenance of Certification (ABMS MOC)[10] standards.

After reviewing existing tools, we concluded that we lacked a tool to assess learners' design and implementation of QI projects. Given that gap, we sought to develop a QI assessment tool that could be used with projects at different stages of development, ranging from proposal to the sustainment phase, and with learners of a variety of training levels—from student through faculty. This article describes the development process for this tool and provides preliminary evidence to support its validity.[11]

Methods

Tool Development

The authors gathered initial content validity evidence by interviewing 4 local QI education leaders on our faculty who had been identified by peers as QI content

experts. These individuals spanned 4 health profession schools (medicine, pharmacy, dentistry, and nursing) and the learner continuum of education (health profession students, graduate students, residents/fellows, and practicing physicians/continuing medical education). In these meetings, 3 investigators (G.R., N.J.B., and S.R.R.) asked these QI content experts/educators to describe the following: (1) what current QI work was performed by their learners; (2) what existing QI curricula were used by the educators; (3) what processes were used to assess QI outputs; and (4) what their specific suggestions or "wishes" were for an ideal assessment tool.

We developed a list of competencies associated with QI work (eg, ability to write an aims statement, identify appropriate measures of change) using information gathered in the meetings. We then cross-referenced that list with materials from the ACGME Milestone Projects,[12] the Society of Hospital Medicine core competencies,[13] as well as project expectations for ABMS MOC requirements[10] to ensure that we did not miss any key competencies. We also drew on our expertise from having designed and led QI curricula for residents in pediatrics (G.R. and N.J.B.) and internal medicine (S.R.R.). We developed an initial list of 6 competency domains; all of which either existed in the commonly accepted frameworks or were highlighted as important by our faculty. The domains were *population*, *stakeholders*, *design*, *measurement*, *evaluation*, and *sustainability.*

This list was then reviewed in subsequent dyad meetings (2 groups, 4 participants total) in which local and national QI educators and administrative leaders (1 of whom had participated in the first round of meetings) were asked to provide unstructured feedback on the domains and to propose any domains that had been omitted. Based on these discussions, 3 additional domains were added to create the final list: (1) *problem identification*, (2) *objective*, (3) *population*, (4) *stakeholders*, (5) *change*, (6) *measures*, (7) *data analysis*, (8) *project evaluation*, and (9) *sustained improvement*.

Our intent was to design a tool in which each item could be scored independently. Although it is assumed that most projects are built sequentially, the nature of learner engagement is such that not all learners have the opportunity to equally participate in all stages of a project. For example, a learner might be assigned to an existing QI project for a limited duration and, therefore, may only be involved in a single test of change but not in population selection or stakeholder identification. Scoring items independently allows for more personalized and customizable feedback to the learners.

What was known and gap
Existing quality improvement (QI) assessment tools have limitations in addressing the range of projects undertaken across the continuum of medical education.

What is new
The Multi-Domain Assessment of Quality Improvement Projects (MAQIP) can be used for formative and summative assessment.

Limitations
The tool cannot fully assess learner engagement or learners' unique contribution.

Bottom line
The MAQIP is a 9-item, user-friendly tool that can be used to assess QI projects by learners at all levels and requires no added faculty training.

Using the list above, we created a 9-item assessment tool, which we called the MAQIP (Multi-Domain Assessment of Quality Improvement Projects; FIGURE). Early iterations used detailed descriptors (anchors) for the 3 levels within each of the 9 items. However, attempts to gather validity evidence revealed that those descriptors, particularly for intermediate levels, were insufficiently discriminating. Subsequently, we developed a single robust descriptor for each item, with 3 levels: *does not fulfill*, *partially fulfills*, and *fulfills*.

We examined the validity evidence for this internal structure with an early focus on interrater reliability. We calculated the intraclass correlation coefficient (ICC) to examine interrater reliability (fair [ICC values 0.21–0.4]; moderate [ICC values 0.41–0.6]; substantial [ICC values 0.61–0.8]) for the global score for each project as well as for each of the 9 items.[14] We piloted the tool by having 3 raters with QI education experience (G.R., S.R.R., N.J.B.) score multiple projects in iterative cycles. The projects were randomly selected from a publicly available library of projects previously completed and presented by students, residents, and faculty at the University of California, San Francisco (UCSF) and the Naval Medical Center, San Diego.

The study was declared exempt by the UCSF Committee on Human Research.

Usability and Acceptability

We conducted a pilot rating session to assess whether faculty raters could use the tool with minimal instruction, as we intended. We invited a convenience sample of UCSF faculty to use an early version of the tool to score projects. We intended the tool to be self-explanatory, and instructions were limited to a brief written notation indicating that it was permissible to skip items the assessor felt did not apply. We provided raters with sample QI projects and asked them to rate

Multi-Domain Assessment of Quality Improvement Projects

Name:

Project:

Problem Identification			Score
Problem is contextualized using evidence. (Evidence may include mandates, guidelines, local gaps in care)			
Does not fulfill 1	Partially fulfills 2	Fulfills 3	

Objective			
Has an AIM statement which fulfills either SMART criteria (specific, measureable, achievable, relevant, time-bound) or all elements of IHI (e.g. who will do what by when)			
Does not fulfill 1	Partially fulfills 2	Fulfills 3	

Population			
Population is appropriate for the QI Project Objective			
Does not fulfill 1	Partially fulfills 2	Fulfills 3	

Stakeholders			
Stakeholders are explicitly identified, evidence of multidisciplinary engagement appropriate to scope and aim			
Does not fulfill 1	Partially fulfills 2	Fulfills 3	

Change			
Systems change or other robust intervention, utilizes an accepted methodology (e.g. PDSA, Six Sigma, Lean)			
Does not fulfill 1	Partially fulfills 2	Fulfills 3	

Measures			
Identifies outcome and/or process measures, as well as at least one balancing measure			
Does not fulfill 1	Partially fulfills 2	Fulfills 3	

Data analysis			
Data are formatted in a run chart with multiple time-points, and include some comment, annotation, or other interpretation of outcome			
Does not fulfill 1	Partially fulfills 2	Fulfills 3	

Project evaluation			
Reflection on initial data with suggestions of next steps			
Does not fulfill 1	Partially fulfills 2	Fulfills 3	

Sustained Improvement			
Demonstrates improvement sustained through multiple cycles and/or time-points			
Does not fulfill 1	Partially fulfills 2	Fulfills 3	

Total Score	

FIGURE
Multi-Domain Assessment of Quality Improvement Projects (MAQIP)

TABLE
Intraclass Correlation for Individual Items in the Multi-Domain Assessment of Quality Improvement Projects

Domains	ICC
1: Problem identification	1.00
2: Objective	0.81
3: Population	0.71
4: Stakeholders	0.53
5: Change	0.50
6: Measures	0.53
7: Data analysis	0.87
8: Project evaluation	0.37
9: Sustained improvement	0.37

Abbreviation: ICC, intraclass correlation coefficient.

these using the tool. Nine raters scored an average of 3 projects each from a sample of 15 projects.

To gather evidence for acceptability by educators and learners, we used the tool to judge and to provide qualitative feedback on projects presented in resident QI symposia at 1 of the other residency programs at our institution over the course of 2 years. Assessors had the opportunity to ask learners clarifying questions, and score adjustments were made based on responses.

Results

The MAQIP was constructed with 9 items corresponding to identified QI competencies and stages of QI projects. We completed 6 rounds of iterative revisions to the language and structure of the tool to achieve acceptable interrater reliability. We scored 14 projects in 2 early rounds to identify language that needed refinement, resolving differences by comparison and discussion. While scoring 28 projects in 3 subsequent rounds, we edited the language in the descriptors to increase discrimination. The project teams included learners of various levels (often in mixed teams), and some teams included faculty. Project formats included abstracts, posters, multipage summaries, and structured online Wiki entries. We calculated interrater reliability after each round using ICC and refined the descriptors based on discrepancies among raters.

We calculated interrater reliability in the sixth (final) round using 10 additional projects that had not been scored previously. Raters used the full range of scores, except for item 3 (only ratings of 2 and 3 were used) and item 4 (only ratings of 1 and 2 were used). Total project scores ranged from 11 to 24 (out of a possible 27). The interrater reliability between 2 raters on the global score for the project was 0.77, and the ICC for individual items ranged from 0.37 to 1.00 (TABLE).

Usability and Acceptability

In the pilot rating session, faculty members did not have questions about using the tool, and all used it as intended. One faculty member who did not have QI experience requested guidance on a technical term. At the resident QI symposium, the education and curriculum leader and the department chair, who used the tool as a judge, perceived it to be appropriate to score projects, select the winner, and guide feedback.

Discussion

The MAQIP tool demonstrated good interrater reliability and acceptability from faculty raters with minimal added training. The instrument is useful for guiding feedback to learners at different levels.

The MAQIP demonstrated internal consistency when used by faculty who were not familiar with the projects they were scoring. Some domains were more difficult to assess with consistency, despite multiple iterative attempts to refine the language. For example, the range of interventions used in projects made it challenging to rate the *change* element, whereas *data analysis* was more straightforward. Ultimately, we achieved acceptable interrater reliability for all items.

Two items (*project evaluation* and *sustained improvement*) had low ICC, attributable to the fact that many projects did not include explanations of these domains because projects were ongoing or that element was not a reporting expectation. We kept these items in the MAQIP because they had been identified by interviewees as important elements of high-quality projects. Raters may choose to use only the elements that apply to their QI learning objectives.

Although we provide preliminary validity evidence for use of the MAQIP as a project assessment tool, the tool can be used in a variety of ways. The domains are useful for guiding learners as they describe key attributes of a QI project. The tool could be given to learners prospectively to help frame key domains and to serve as a reference as well as a guide for feedback later. Educators might use certain items for summative assessment and other items as a guide for formative feedback, based on specific goals and objectives of the given educational experience.

Limitations of the tool are that assessment relies on a retrospective description of the QI project and that the tool cannot directly assess learners' unique

contributions during project execution, an element that had been requested by several experts. We also attempted to quantify the level of learner engagement in projects but found it difficult with the data available, and thus, the tool does not account for the possibility of varying levels of involvement. We also primarily used archived projects for tool development and were not able to assess learner perceptions of the tool. As we continue to use the tool, this will be an important assessment dimension.

Future work with the MAQIP should focus on collecting additional evidence for validity, including usability both to prospectively guide QI projects and to retrospectively assess QI projects. It also would be useful to know if the MAQIP correlates with other tools, such as learners' knowledge assessments and project assessments. Finally, future work should explore metrics to assess individual engagement in QI work, including self-reflection, peer assessment, and active discussion with participants.

Conclusion

We developed the MAQIP tool to assess the quality of QI projects at all levels of medical education. The tool has built-in flexibility to use all or part of the 9 domains by faculty, without requiring added faculty training in use of the instrument. The MAQIP demonstrates acceptable interrater reliability, and is suitable for use in rating projects as well as providing feedback to learners.

References

1. Headrick LA, Baron RB, Pingleton SK, et al. Teaching for Quality: Integrating Quality Improvement and Patient Safety Across the Continuum of Medical Education. Association of American Medical Colleges. January 2013. https://members.aamc.org/eweb/upload/Teaching%20for%20Quality%20Report.pdf. Accessed May 26, 2017.
2. Accreditation Council for Graduate Medical Education. Common Program Requirements. http://www.acgme.org/What-We-Do/Accreditation/Common-Program-Requirements. Accessed July 10, 2017.
3. Accreditation Council for Graduate Medical Education. *CLER Pathways to Excellence*. http://www.acgme.org/acgmeweb/Portals/0/PDFs/CLER/CLER_Brochure.pdf. Accessed May 26, 2017.
4. Weiss KB, Bagian JP; CLER Evaluation Committee. Challenges and opportunities in the six focus areas: CLER national report of findings 2016. *J Grad Med Educ*. 2016;8(2 suppl 1):25–34.
5. Ogrinc G, Headrick LA, Morrison LJ, et al. Teaching and assessing resident competence in practice based learning and improvement. *J Gen Intern Med*. 2004;19(5, pt 2):496–500.
6. Singh MK, Ogrinc G, Cox KR, et al. The quality improvement knowledge application tool revised (QIKAT-R). *Acad Med*. 2014;89(10):1386–1391.
7. 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.
8. Wittich CM, Beckman TJ, Drefahl MM, et al. Validation of a method to measure resident doctors' reflections on quality improvement. *Med Educ*. 2007;44(3):248–255.
9. 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.
10. American Board of Medical Specialties. *Based on Core Competencies*. http://www.abms.org/Maintenance_of_Certification/MOC_competencies.aspx. Accessed May 26, 2017.
11. Messick S. Validity. In: Linn RL, ed. *Educational Measurement*. 3rd ed. New York, NY: MacMillan; 1989:13–103.
12. Accreditation Council for Graduate Medical Education. *Milestones*. http://www.acgme.org/What-We-Do/Accreditation/Milestones/Overview. Accessed May 26, 2017.
13. Society of Hospital Medicine. Quality improvement. *J Hosp Med*. 2006;1(suppl 1):92. http://onlinelibrary.wiley.com/doi/10.1002/jhm.61/full. Accessed May 26, 2017.
14. Fleiss JL, Shrout PE. Approximate interval estimation for a certain intraclass correlation coefficient. *Psychometrika*. 1978;43(2):259–262.

Glenn Rosenbluth, MD, is Associate Clinical Professor of Pediatrics, Director of Quality and Safety Programs, Office of Graduate Medical Education, and Associate Director, Pediatrics Residency Training Program, University of California, San Francisco; **Natalie J. Burman, DO, MEd,** is Assistant Professor of Pediatrics, Uniformed Services University of Health Sciences, and Associate Pediatric Program Director, Naval Medical Center, San Diego; **Sumant R. Ranji, MD,** is Professor of Medicine, University of California, San Francisco, and Chief of the Division of Hospital Medicine, Zuckerberg San Francisco General Hospital; and **Christy K. Boscardin, PhD,** is Associate Professor of Medicine, and Education Faculty of the Office of Medical Education, University of California, San Francisco.

Funding: This study was funded with an Innovations Funding for Education Grant through the Haile T. Debas Academy of Medical Educators, University of California, San Francisco.

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

An earlier version of this work was presented as a poster at the Association of American Medical Colleges Annual Meeting,

Philadelphia, Pennsylvania, November 1–6, 2013, and on MedEdPORTAL's iCollaborative Web site (https://www.mededportal.org/icollaborative/resource/841).

The authors would like to thank Robert Baron, MD, MS, and Andrew Auerbach, MD, MPH, for early feedback on the tool, as well as Weston Fisher, MD, and Matthew State, MD, PhD, who invited us to use the tool in their resident quality improvement symposium.

Corresponding author: Glenn Rosenbluth, MD, University of California, San Francisco, Department of Pediatrics, 5th floor, Mission Hall Box 3214, 550 16th Street, San Francisco, CA 94143, 415.476.9185, glenn.rosenbluth@ucsf.edu

Received January 17, 2017; revision received April 6, 2017; accepted April 26, 2017.