INNOVATIONS IN EDUCATION

Validation of a Method for Assessing Resident Physicians' Quality Improvement Proposals

James L. Leenstra, MD¹, Thomas J. Beckman, MD², Darcy A. Reed, MD, MPH³, William C. Mundell, MD², Kris G. Thomas, MD³, Bryan J. Krajccek, MD¹, Stephen S. Cha, MS⁴, Joseph C. Kolars, MD⁵, and Furman S. McDonald, MD MPH²

¹Division of General Internal Medicine, Mayo Clinic College of Medicine, Rochester, MN 55905, USA; ²Division of General Internal Medicine, Mayo Clinic College of Medicine, Rochester, MN 55905, USA; ³Division of Primary Care Internal Medicine, Mayo Clinic College of Medicine, Rochester, MN 55905, USA; ⁴Division of Biostatistics, Mayo Clinic College of Medicine, Rochester, MN 55905, USA; ⁵Division of Gastroenterology and Hepatology, Mayo Clinic College of Medicine, Rochester, MN 55905, USA.

BACKGROUND: Residency programs involve trainees in quality improvement (QI) projects to evaluate competency in systems-based practice and practice-based learning and improvement. Valid approaches to assess QI proposals are lacking.

OBJECTIVE: We developed an instrument for assessing resident QI proposals—the Quality Improvement Proposal Assessment Tool (QIPAT-7)—and determined its validity and reliability.

DESIGN: QIPAT-7 content was initially obtained from a national panel of QI experts. Through an iterative process, the instrument was refined, pilot-tested, and revised.

PARTICIPANTS: Seven raters used the instrument to assess 45 resident QI proposals.

MEASUREMENTS: Principal factor analysis was used to explore the dimensionality of instrument scores. Cronbach's alpha and intraclass correlations were calculated to determine internal consistency and interrater reliability, respectively.

RESULTS: QIPAT-7 items comprised a single factor (eigenvalue=3.4) suggesting a single assessment dimension. Interrater reliability for each item (range 0.79 to 0.93) and internal consistency reliability among the items (Cronbach's alpha=0.87) were high.

CONCLUSIONS: This method for assessing resident physician QI proposals is supported by content and internal structure validity evidence. QIPAT-7 is a useful tool for assessing resident QI proposals. Future research should determine the reliability of QIPAT-7 scores in other residency and fellowship training pro-

grams. Correlations should also be made between assessment scores and criteria for QI proposal success such as implementation of QI proposals, resident scholarly productivity, and improved patient outcomes.

KEY WORDS: quality improvement; systems-based practice; practice-based learning and improvement; assessment; evaluation study; validation study.

DOI: 10.1007/s11606-007-0260-y

© 2007 Society of General Internal Medicine 2007;22:1330–1334

INTRODUCTION

The Accreditation Council for Graduate Medical Education (ACGME) released its Outcome Project in 1999 requiring residents and fellows to demonstrate competency in systems-based practice (SBP) and practice-based learning and improvement (PBLI).¹ Involvement of trainees in quality improvement (QI) efforts has been proposed as a means of addressing these two competencies.

Indeed, resident physicians are in an excellent position to identify health care systems improvements,^{2–4} yet they often lack the skills to identify the best interventions for improvement and the time and mentorship necessary to implement them. Consequently, many internal medicine residencies have developed QI curricula. We searched the literature for examples of health systems and QI curricula and their assessment.^{3,5–10} Some published curricula involve seminar and lecture formats that are typically evaluated using pre- and posttests of knowledge.^{5,7,8,10} Other curricula incorporate independent-study QI proposals and/or QI projects.^{3,5–10} However, reports of QI project assessments, and QI proposals in particular, are uncommon. Like any thoughtful research or business endeavor, a good QI project begins with a written proposal that explains the problem, hypothesizes a best solution, and outlines a plan to evaluate outcomes. However, we are unaware of any methods for evaluating such QI proposals. Therefore, we created a QI proposal assessment tool and determined the validity of its scores for evaluating resident QI proposals.

Study findings were presented at the Association of Medical Education in Europe Annual Meeting, Genoa, Italy, September 14–18, 2006 and at the ABMS/ACGME Joint Conference on Assessing and Improving Patient Care, Rosemont, IL, November 2–3, 2006.

Received December 20, 2006

Revised May 3, 2007

Accepted June 4, 2007

Published online June 30, 2007

METHODS

Quality Improvement Curriculum

The Mayo Clinic Internal Medicine Residency has 144 categorical residents including approximately 48 residents in each of 3 years of training. A major goal of the program is to train residents in quality improvement for which a curriculum has been developed over the last several years. The content and instructional methods of the curriculum were initially based on Achieving Competency Today (ACT), a national initiative of the Robert Wood Johnson Foundation, administered by Partners for Quality Education, to develop curricula for graduate medical education in the ACGME competencies of SBP and PBLI.¹¹ Our residency program was 1 of 17 programs participating in this initiative between 2003 and 2005.

In the Mayo curriculum, each month, 8 residents work in pairs of one first year and one third year resident to complete a QI proposal. These proposals are based on principles outlined in the curriculum, which incorporates recommendations from the ACT consensus group (Table 1) and the Plan-Do-Study-Act (PDSA) model of QI that was adapted to health care from the business setting.¹²⁻¹⁴ Residents are encouraged to develop a habit of identifying patient care-related QI opportunities encountered in ambulatory and hospital settings. Residents then measure the magnitude of the problem, consult key stakeholders, identify root causes, prioritize possible interventions based on an analysis of likely effort and yield, and detail a plan for implementation and assessment of the intervention. In total, each resident dedicates 64 h to this curriculum during a 1-mo ambulatory rotation in the first and third year of residency. They receive weekly guidance from chief residents and faculty mentors. At rotation end, residents submit written QI proposals for evaluation. They also present their proposals to the faculty and institutional QI leadership to exchange ideas and receive formative feedback intended to help them with future implementation of the QI proposal. Not all resident QI proposals have been implemented, and the implementation rate is being monitored for the purpose of future study.

QI Proposal Assessment Tool

Content of our QI proposal assessment tool was established by input from national and Mayo experts. Mayo experts (the study authors) were experienced in teaching the QI curriculum and mentoring resident QI projects, and one author (TJB) had experience with psychometric assessment and scale design in medical education. In May 2004, the Second Annual ACT Conference convened a session during which faculty participants identified the necessary components of a QI proposal (see Table 1). On the basis of this information, we created an initial assessment instrument.

The initial assessment instrument was tested by faculty and residents on a monthly basis, as part of the curricular sessions described above, from July to November 2005. Faculty and residents openly discussed their assigned ratings and opinions regarding the instrument's usability. The items were modified based on these discussions. The initial instrument had 9 items. Three of these items—"timeline for implementation", "assessment of the intervention", and

Table 1. Consensus on the Necessary Components and Assessment Methodology of Quality Improvement Proposals from the Second Annual Achieving Competence Today Conference, May 2004

Components of the QI proposal
1. Represents a problem of merit
2. Includes a well-defined problem
3. Includes a statement describing how change can be measured
4. Shows evidence of a root cause analysis having been conducted
5. Suggests an intervention that directly addresses the problem
6. Suggests a change that involves all stakeholders in terms of (a) identifying who the stakeholders are, (b) including them in the analysis, and (c) interviewing them
7. Is of an appropriate scope given time and resources
8. Involves cooperation among key personnel
Assessment of the QI proposal
Issues of reliability may occur when multiple raters are used. Clearly, the raters will need some training so that they use a common set of standards. The following ideas were generated to improve reliability:
1. Use <i>multiple raters</i> appropriately, asking them to assess only those aspects of the QI proposal they have expertise in (e.g., many patients will not know how to conduct a root cause analysis but will be able to testify to their perception of the value of an intervention). Raters might include the ACT faculty, other residents who hear a presentation of a QI proposal, QI experts (e.g., external consultants), other health care providers, and the resident him/herself
2. Use a Likert scale with <i>anchoring terms</i> to define clearly what is/is not satisfactory
3. Provide a <i>gold standard</i> report so that everyone has an understanding of what goes into a QI proposal and what a good idea looks like and why it is a good idea
4. Use a <i>cadre of assessors</i> who come in to hear or read QI proposals periodically. They should understand the QI process as well as be able to calibrate good and unsatisfactory performance among residents
5. Use <i>criteria familiar</i> to raters from work in assessing performance in other domains

"scope of the project"—were modified by the group. The first two were condensed into the item "implementation and evaluation of the intervention", and the latter was eliminated. Through this iterative process, we developed the 7-item, Quality Improvement Proposal Assessment Tool (QIPAT-7, Fig. 1). The items, structured on a 5-point rating scale (1=needs improvement, 2 to 4=meets expectations, and 5=exceeds expectations), included the following: definition of the problem, identification of key stakeholders, root cause analysis, choice of the QI project, identification of potential interventions, proposed intervention, and plan for implementation and evaluation of the intervention.

After approval by the Mayo Clinic Institutional Review Board, QIPAT-7 was pilot-tested in December of 2005 by 5 faculty members and 2 chief medical residents (all study authors except TJB and SSC), who individually rated 3 randomly chosen resident QI proposals from the previous academic year. These raters then met and, after critical discussion, resolved all differences in their assigned ratings. On the basis of this discussion, it was decided that each item's anchor descriptions (Fig. 1) should be achieved to warrant a score of 3 or more for that item. From January to March 2006, the 7 raters used QIPAT-7 to score 45 consecutive resident QI proposals that had been completed from July 2004 through July 2005. The data collection was accomplished using secure electronic forms and computerized scoring. Data from these ratings were used in the current study.

Data Analysis

Exploratory principal factor analysis was used to determine whether items clustered into separate domains. The eigenvalue rule (retaining factors with eigenvalues >1) was used to identify factors. Results of this method were confirmed by inspecting the corresponding scree plot.^{15,16} For each item, scores assigned by the raters were reported as means and standard deviations. Interrater reliability was established by calculating intraclass

correlation coefficients with 95% confidence intervals. Intraclass correlation coefficients were interpreted as follows: <0.4=poor agreement, 0.4 to 0.75=fair to good agreement, and >0.75=excellent agreement.¹⁷ Internal consistency reliability was determined by calculating Cronbach's coefficient alpha. By convention, alpha >0.7 was considered acceptable. All calculations were performed using SAS (version 9.1.3, SAS Institute, Cary, NC).

All bullets must be satisfied in each domain for a rating of 3 or higher

Definition of the problem <ul style="list-style-type: none"> Establishes problem magnitude/significance Identifies affected groups Clear statement of the problem 	Needs improvement	Meets expectations			Exceeds expectations
	1	2	3	4	5
	☐	☐	☐	☐	☐
	Comment				
Identification of key stakeholders <ul style="list-style-type: none"> Evidence of stakeholder consultation Description of impact of proposed intervention on stakeholders 	Needs improvement	Meets expectations			Exceeds expectations
	1	2	3	4	5
	☐	☐	☐	☐	☐
	Comment				
Evidence of root cause analysis <ul style="list-style-type: none"> Prioritizes causal factors Identified systems issues Utilizes at least one QI tool (eg, fishbone, systems walk, mind map) 	Needs improvement	Meets expectations			Exceeds expectations
	1	2	3	4	5
	☐	☐	☐	☐	☐
	Comment				
Choice of quality improvement project <ul style="list-style-type: none"> Likely to result in meaningful improvement to patient care (eg, clinical outcomes, safety, efficiency, or cost) Stimulates further inquiry 	Needs improvement	Meets expectations			Exceeds expectations
	1	2	3	4	5
	☐	☐	☐	☐	☐
	Comment				
Potential interventions <ul style="list-style-type: none"> Prioritization of multiple interventions Effort vs yield analysis 	Needs improvement	Meets expectations			Exceeds expectations
	1	2	3	4	5
	☐	☐	☐	☐	☐
	Comment				
Proposed intervention <ul style="list-style-type: none"> Directly addresses the problem Reasonable potential to change system for the better Impact on care captures ≥ 2 of the following (high yield/low effort, innovative, cost effective, sustainable) 	Needs improvement	Meets expectations			Exceeds expectations
	1	2	3	4	5
	☐	☐	☐	☐	☐
	Comment				
Implementation and evaluation of the intervention <ul style="list-style-type: none"> Clear plan and timeline for implementing the intervention Identifies measures of intervention success/effectiveness 	Needs improvement	Meets expectations			Exceeds expectations
	1	2	3	4	5
	☐	☐	☐	☐	☐
	Comment				
Summary comments					

Figure 1. Quality Improvement Proposal Assessment Tool (QIPAT-7). The scale is anchored to the bulleted comments on the left. To achieve a score of 3 or higher, all bullets for each domain must be met. The box sizes for each point of the scale are simply determined by the heading labels; that is, smaller boxes do not indicate smaller intervals between scale steps.

Table 2. Quality Improvement Proposal Assessment Tool: Mean Scores and Reliability

Items	Mean	S.D.	ICC*	95% CI	Cronbach's $\alpha^†$
Definition of the problem	3.4	0.995	0.79	0.61, 0.84	0.87
Identification of stakeholders	2.6	1.16	0.91	0.84, 0.94	
Root cause analysis	2.3	1.05	0.93	0.87, 0.95	
Choice of QI project	3.4	1.08	0.82	0.62, 0.84	
Potential interventions	2.1	0.968	0.87	0.78, 0.91	
Proposed intervention	2.9	1.1	0.85	0.65, 0.86	
Implementation and evaluation	1.9	1.12	0.90	0.82, 0.93	

*ICC = Intraclass correlation coefficient, a measure of interrater reliability

†Cronbach's alpha represents internal consistency reliability for the 7 items.

RESULTS

Principal factor analysis revealed that the items comprised a single factor (eigenvalue=3.4), indicating that all 7 items represent 1 dimension of QI proposal assessment. This factorial model accounted for nearly 100% of the shared variance among the original variables. Item mean scores ranged from 1.9 to 3.4 on a 5-point scale. Interrater reliability for each item (range 0.79 to 0.93) and internal consistency reliability among the items (Cronbach's alpha=0.87) were excellent (Table 2).

DISCUSSION

With the introduction of the competencies of SBP and PBLI, the ACGME recognized that resident physicians require specific training in quality improvement.¹ Phase 2 of the ACGME Outcome Project aimed to improve resident assessment in all of the ACGME competencies.¹⁸ Because SBP and PBLI were new concepts to most educators,⁹ there are very few assessment methods for their domains.¹⁹ Nonetheless, the ACGME Outcome Project asserts that "... programs are expected to phase-in assessment tools that provide useful and increasingly valid, reliable evidence that residents achieve competency-based educational objectives."¹ Here we report the first valid and reliable method for assessing resident QI proposals created to meet the SBP and PBLI core competencies.

Meaningful assessment measures must be reliable and valid.²⁰ Current theory states that all validity is construct validity, and evidence is collected to support validity from the following 5 sources: content, response process, internal structure, relations to other variables, and consequences.²⁰⁻²³ Recent reviews of published clinical teaching assessments revealed that the most commonly reported and highly valued categories of validity evidence are those of content and internal structure, which include reliability.^{24,25} Clearly, assessments will be valid to the extent that their content accurately reflects the constructs of interest, which in this case are SBP and PBLI. Likewise, validity will be enhanced to the extent that raters agree (are reliable) on their assessments of SBP- and PBLI-related proposals and projects.

Despite the ACGME call for the valid assessments of competency-based residency curricula,¹ we are unaware of any previous, standardized methods for assessing resident physician QI proposals. We report strong content and internal structure validity evidence for scores from our instrument, QIPAT-7. Internal structure validity evidence was supported by an assessment of the instrument's dimensionality (factor analysis) and high internal consistency and interrater reliability. Content validity evidence was supported by input from a panel of national experts, and instrument refinement through the use of a careful, prospectively planned iterative process. Response process validity evidence was supported, in part, by the use of secure electronic forms and computerized scoring.

Although we found no previous, reliable methods for assessing resident QI proposals, we acknowledge existing recommendations regarding formats for reporting of QI projects. Moss and Thompson²⁶ originally suggested that published QI studies contain the following elements: brief description of the context, outline of the problem, key measures for improvement, process of gathering information, analysis and interpretation, strategy for change, effects of change, and next steps. Notably, each of Moss and Thompson's suggested elements is represented in QIPAT-7. Davidoff and Batalden expanded on this structure by organizing 16 similar items into the IMRAD (introduction, methods, results, and discussion) format.²⁷ Our assessment method incorporates these previously published elements of QI project reporting and, to the extent that assessment drives learning and performance, it should facilitate the effective communication of QI study results.

We also recognize previous assessments of QI curricula that are structured on lecture and seminar formats. For example, Ogrinc et al.⁸ described an assessment of 11 resident QI project presentations. Faculty members rated these presentations using the following scaled items (1=unsatisfactory to 4=excellent): developing an aim, describing and diagramming the work process, linking data to change, describing the PDSA cycle, eliciting interdisciplinary input, and incorporating interdisciplinary perspective. Once again, these assessment elements are strongly represented in QIPAT-7. However, psychometric performance of the data of Ogrinc et al. is lacking, and their assessment targeted project presentations, not written QI proposals.

Our study has several advantages. Content validity evidence is strongly supported by compatibility with published recommendations for QI reporting and assessment,^{8,26,27} national expert consensus, and instrument refinement through a careful, iterative process. We anticipate that QIPAT-7's content will have broad appeal to other residency programs, many of which likely have or will develop curricula that require residents to complete QI proposals and projects. Additionally, QIPAT-7 scores have excellent internal consistency and interrater reliability, which are essential elements of valid assessments.²⁴ We also found that utilization of QIPAT-7 is feasible in a large internal medicine residency program, as single ratings could be completed within 15 min, including the time required to read the proposal. Finally, we found that critical application of QIPAT-7 enhanced raters' understanding of quality improvement itself.

Our study has limitations. Our findings are based on results from a single institution. Nonetheless, QIPAT-7 content was derived from a panel of national experts from many institutions, and despite several iterative revisions (which are recog-

nized as essential for scale development), the final instrument retained content recommended by the original panel of national experts (see Table 1 and Fig. 1). Hence, our assessment method may generalize to other settings. Another limitation is that we describe the assessment of a time-consuming curriculum. Indeed, Djuricich et al.⁷ who describe a curriculum very similar to ours highlight the difficulties of designing a resident QI project in 1 mo, let alone completing the project. All the same, it is necessary for residents to carefully plan any project that will eventually be successfully implemented, and meaningful assessment is an inseparable component of formative feedback on QI proposals. Yet another limitation is that our instrument was shown to be reliable among a group of raters who were very familiar with QI concepts and were integral to the scale's development, and for a QI curriculum that may have elements unique to Mayo Clinic's learning environment. Therefore, the reliability of our instrument's scores and the generalizability of our findings must be considered cautiously and in this context. Finally, although our sample size was adequate for an accurate factor analysis, we recognize that repeating this study on a much larger sample may show that QIPAT-7 captures more than 1 dimension of QI proposal assessment.

We suggest that future studies replicate QIPAT-7 in other educational settings or correlate scores from QIPAT-7 and other QI assessment instruments. Moreover, we plan to seek "relations to other variables" validity evidence^{20,22} by determining whether QIPAT-7 scores correlate with other meaningful criteria, such as selection of resident QI projects for presentation and publication, actual implementation of the project, or improved patient outcomes. With the accumulation of study findings on QIPAT-7 from other institutions and additional validity evidence, this method may prove useful for assessing resident QI proposals and projects.

Acknowledgments: We are grateful to Gordon T. Moore, M.D., M.P.H., Maryjoan Ladden, Ph.D., R.N., C.S., Antoinette Peters, Ph.D., and the faculty of the other academic medical centers involved in the Achieving Competency Today initiative for their input at the original forum discussing content that formed the basis for the first iteration of QIPAT-7. We also extend our thanks to Mr. Gregory J. Engstler and Ms. Pamela J. Nelson for their technical assistance with the computerized version of the assessment tool.

There were no external or internal funding sources for this work.

Conflict of Interest: None of the authors has any conflict of interest related to this work.

Corresponding Author: Furman S. McDonald, MD, MPH; Division of General Internal Medicine, Mayo Clinic College of Medicine, Rochester, MN 55905, USA (e-mail: mcdonald.furman@mayo.edu).

REFERENCES

1. ACGME Outcome Project. Available at: <http://www.acgme.org/outcome/comp/compFull.asp>. Accessed July 29, 2006.

2. Ashton CM. "Invisible" doctors: making a case for involving medical residents in hospital quality improvement programs. *Acad Med*. 1993;68(11):823-4.
3. Headrick LA, Richardson A, Priebe GP. Continuous improvement learning for residents. *Pediatrics*. 1998;101(4 Pt 2):768-73; discussion 773-4.
4. Parenti CM, Lederle FA, Impola CL, Peterson LR. Reduction of unnecessary intravenous catheter use. Internal medicine house staff participate in a successful quality improvement project. *Arch Intern Med*. 1994;154(16):1829-32.
5. Frey K, Edwards F, Altman K, Spahr N, Gorman RS. The 'Collaborative Care' curriculum: an educational model addressing key ACGME core competencies in primary care residency training [see comment]. *Med Educ*. 2003;37(9):786-9.
6. Amin AN, Rucker L. A systems-based practice curriculum. *Med Educ*. 2004;38(5):568-9.
7. Djuricich AM, Ciccirelli M, Swigonski NL. A continuous quality improvement curriculum for residents: addressing core competency, improving systems. *Acad Med*. 2004;79(10 Suppl):S65-7.
8. 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.
9. Ziegelstein RC, Fiebach NH. "The mirror" and "the village": a new method for teaching practice-based learning and improvement and systems-based practice. *Acad Med*. 2004;79(11):83-8.
10. Allen E, Zerzan J, Choo C, Shenson D, Saha S. Teaching systems-based practice to residents by using independent study projects. *Acad Med*. 2005;80(2):125-8.
11. ACT (Achieving Competence Today) Online. Available at: <http://www.actcurriculum.org/>. Accessed June 27, 2006.
12. Langley GJ, Nolan KM, Nolan TW, Norman CL, Provost LP. The Improvement Guide: A Practical Approach to Enhancing Organizational Performance. San Francisco, CA: Jossey-Bass; 1996.
13. Cleghorn GD, Headrick LA. The PDSA cycle at the core of learning in health professions education. *Joint Comm J Qual Improv*. 1996;22(3):206-12.
14. Berwick DM, Nolan TW. Physicians as leaders in improving health care: a new series in *Annals of Internal Medicine* [see comment]. *Ann Intern Med*. 1998;128(4):289-92.
15. DeVellis RF. Scale Development: Theory and Applications. London: Sage Publications; 1991.
16. Gorsuch RL. Factor Analysis, Second Edition. Hillsdale: Lawrence Erlbaum Associates, Inc.; 1983.
17. Landis JR, Koch GG. The measure of observer agreement for categorical data. *Biometrics*. 1977;33:159-74.
18. Swing SR. Assessing the ACGME general competencies: general considerations and assessment methods. *Acad Emerg Med*. 2002;9(11):1278-88.
19. Lynch DC, Swing SR, Horowitz SD, Holt K, Messer JV. Assessing practice-based learning and improvement. *Teach Learn Med*. 2004;16(1):85-92.
20. Messick S. Validity. In: Linn RL, ed. Educational Measurement. 3rd ed. Phoenix, AZ: Oryx Press; 1993.
21. Downing SM. Validity: on the meaningful interpretation of assessment data. *Med Educ*. 2003;37:830-7.
22. Standards for Educational and Psychological Testing. Washington, DC: American Education Research Association; 1999.
23. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. *Am J Med*. 2006;119:166e7-16.
24. Beckman TJ, Ghosh AK, Cook DA, Erwin PJ, Mandrekar JN. How reliable are assessments of clinical teaching? A review of the published instruments. *J Gen Intern Med*. 2004;19:971-7.
25. Beckman TJ, Cook DA, Mandrekar JN. What is the validity evidence for assessments of clinical teaching? *J Gen Intern Med*. 2005;20:1159-64.
26. Moss F, Thompson R. A new structure for quality improvement reports [comment]. *Qual Health Care*. 1999;8(2):76.
27. Davidoff F, Batalden P. Toward stronger evidence on quality improvement. Draft publication guidelines: the beginning of a consensus project [see comment]. *Qual Saf Health Care*. 2005;14(5):319-25.