

The evaluation of quality assurance: developing and testing practical methods for managers

TESSIE TZAVARAS CATSAMBAS¹, ED KELLEY², STEPHANE LEGROS³, RASHAD MASSOUD⁴ AND BRUNO BOUCHET⁵

¹EnCompass LLC, Bethesda, MD, ²Agency for Health Research and Quality, Rockville, MD, ³JHPIEGO, Johns Hopkins University, Baltimore, MD, ⁴Quality Assurance Project, University Research Co., LLC, Bethesda, MD, ⁵John Snow, Inc., Boston, MA, USA

Abstract

Objective of the study. To develop an approach for evaluating quality assurance (QA) activities and programs in health care settings and to test different evaluation methods.

Design. This was not a formal scientific study, but rather a research and development (R&D) study, which followed the following steps: (1) reviewing the literature; (2) clarifying critical issues for all key aspects of QA activities; (3) drafting a guide to provide a flexible vehicle for different approaches; (4) testing and adapting the guide as it evolved in three countries; and (5) testing two evaluation tools (self-assessment and appreciative evaluation) in Chile.

Setting and study participants. The evaluation guide was tested by evaluating QA structures, activities, and programs at the country, regional, and facility levels in Zambia, Niger, and Chile.

Results. The study resulted in an evaluation guide, which includes an implementation outline, an evaluation matrix, and an appendix of evaluation tools and methods. The guide helps evaluators: agree on a proposed evaluation's scope and design; develop an evaluation methods plan; and address QA history, advocacy, culture, and structure, as well as QA activities and accomplishments. Specific results of the country evaluations in Zambia, Niger, and Chile are presented in separate articles in this supplement.

Conclusion. The QA programs in which the evaluation guide was tested differed in many ways, such as health system structure, decision to focus on particular services, political level implementing QA, policy environment, leadership, and program evolution. The implementation guide presents an outline of the key implementation steps for an evaluation, and includes checklists and model forms (e.g. sample agenda for a team planning meeting, sample list of questions to focus the evaluation). The evaluation matrix presents indicators by QA component and key question, and it enables evaluators to build an approach and select methods. The appendix describes the various tools and methods presented in the first two sections of the evaluation guide.

Keywords: appreciative inquiry, Chile, evaluation, Niger, self-assessment, Zambia

The United States Agency for International Development (USAID)-funded Quality Assurance Project II (QAP) required an approach and testing methods to evaluate QA activities and programs. As the work of designing evaluation methods was underway, USAID missions asked QAP to evaluate selected QA programs. This provided the opportunity to review different approaches interactively with field visits, to build an evaluation guide, and to field-test

various methods. QAP evaluators teamed up with outside experts to conduct three evaluations, in Zambia, Niger and Chile (each is presented in separate articles in this supplement). QAP technical staff engaged in probing discussions about the 'core QA approach' versus 'field adaptations', and the implications of this for how to evaluate activities and programs. Rather than develop one approach for evaluating QA, the team developed an evaluation guide that would allow

Address reprint requests to Tessie Tzavaras Catsambas, EnCompass LLC, 7803 Fox Gate Court, Bethesda, MD 20817, USA. E-mail: tcatsambas@encompassworld.com
E. Kelley, S. Legros, and B. Bouchet were formerly with the Quality Assurance Project, University Research Co., LLC, Bethesda, MD, USA.

evaluators to build an approach appropriate for each country and setting, and select appropriate methods.

Study objective

The objective of this research and development (R&D) study was to develop an approach for evaluating QA activities and programs in health care settings, and to test different evaluation methods.

Methods

The evolution of this R&D effort included the following steps/phases:

- (1) Prior to this study, QAP had developed a QA evaluation matrix based on an input–output model (1977) that provided a starting point.
- (2) Concurrent with the initiation of this R&D effort, QAP sent an evaluation team to Zambia (1998). The team used a systems guide for organizing its work and presenting evaluation results. This was found to be useful, and the evaluation showed that before any specific methods were tested, a clear list of evaluation questions was needed for each QA activity. Furthermore, the Zambia team recommended that an implementation outline be developed, because success in an evaluation is due as much to sound methodology as it is to well managed client negotiations, team relations, and logistics. Lastly, the Zambia evaluation tested a ‘scoring’ approach to the evaluation, over which the team itself was divided and which was not well received by the client. This raised the questions of whether the R&D study should present standards and thresholds for evaluating QA, or whether a more flexible approach was needed.
- (3) The 1997 input–output matrix, informed by the Zambia systems model, evolved into a multipage outline organized by QA activity, including key evaluation questions for each activity (1998–99). This was presented and endorsed by QAP staff and the USAID Project Officer (July 1999).
- (4) The Niger evaluation (early 1999) followed an approach that compared progress against plan. The Niger team suggested that the then current evaluation outline was insufficient. They needed a tool with more guidance (e.g. indicators, samples, methods) that would help them move faster in the field. Also, the Niger team indicated that evaluators needed clarity on what to do prior to the visit. This led to the transformation of the multipage outline into an evaluation matrix, essentially an expansion of the 1997 evaluation matrix.
- (5) The evaluation matrix was used in the Chile evaluation (1999). Having finally agreed on the first two parts of the guide, the R&D team was ready to test specific evaluation tools. In Chile, two tools were tested: self-assessment and appreciative evaluation.

Results

Overview

Rather than develop one approach, this R&D study developed a flexible guide that gives evaluators options for crafting an evaluation approach, and methods in ways that accommodate differences between countries, districts, and facilities. Furthermore, the guide enables evaluators to include in their scope both formal QA program activities and other activities that may not be formally labeled as QA or quality improvement, but actually contribute to assuring quality in a health care system. The guide is intended for use by internal or external evaluators of developing country QA programs at any level of the health system (e.g. national, regional). It is not intended for assessing the quality of health care or as a substitute for accreditation. It has three components:

Part one

The Quality Assurance Implementation Outline leads managers and evaluators step-by-step through the key issues and decisions that must be made in planning and conducting an evaluation. It also provides checklists and highlights decision steps to help evaluators develop their own methods and implementation plans.

Part two

The Quality Assurance Evaluation Matrix is a 30-page table organized by QA activity with a menu of key questions and issues; it also suggests key indicators to measure, data sources, and a data collection schedule.

Part three

Evaluation Tools and Methods is a compendium of suggested evaluation tools, the most appropriate use for each, and instructions for how to apply them.

Part one (implementation outline) explains the matrix and offers step-by-step basic guidance for implementing an evaluation. Part two (matrix) was developed as a menu of questions and indicators to help plan and focus the evaluation. Part three (tools) was developed as a reference describing the tools and methods discussed in the first two parts. These components grew from a recognition that this product had to serve several objectives: (1) to develop an understanding of the nature and salient features of QA activities and programs to focus the evaluation; (2) to raise practical issues related to managing an evaluation to ensure its success; and (3) to clarify and illustrate how to apply the methods and tools.

Challenges

Two challenges emerged at the outset: (1) how to categorize and evaluate different QA organizational structures; and (2) how to ensure consistency in evaluating QA programs, possibly by ‘grading’ or otherwise measuring the performance of any given QA program.

1. Categorizing and evaluating QA structures

Two opposite positions were considered: (1) that QA organizational structure is critical and its absence should be identified as a gap; and (2) that QA organizational structure is only one element and not necessarily of primary importance. Ultimately, the issue was reframed: rather than evaluating organizational structure itself, we concluded that understanding the history and evolution of the program, as well as its current structure, was important in framing the rest of the evaluation. Consequently, we developed a QA program map that would provide historical milestones, and capture and document various dimensions of QA activities and programs: emphasis on QA activities (e.g. training, standards, QA improvement), selection of services where QA activities are focused (e.g. tuberculosis, acute respiratory infection), level where QA activities are implemented (national, district, local, facility), and locus of decision making, leadership, culture and so forth.

2. Ensuring consistency between evaluations

As might be expected, it was not possible to agree on the 'right way' to approach and implement QA, making it even harder to develop a quantitative evaluation system. Firstly, countries vary widely in every aspect: existing health care system, resources, advocacy, training, infrastructure, policy environment, and QA program evolution, structure, and focus. Secondly, who had the 'authority' to define the 'right' way? Finally, we had concerns about who would be served by an evaluation that 'grades' QA activities and what effect such grading would have on those implementing QA activities. We were challenged to develop an evaluation guide that would offer a level of consistency that would make the evaluation results useful for improving QA activity or program performance. Consistency was difficult, partly due to evaluators' different preconceptions and also due to differences in evaluation objectives, which also influenced choices about evaluation plans and methods. Consequently, the Zambia, Niger, and Chile evaluations differed in part because of the team composition and different client needs. These differences are described as follows:

The Zambia evaluation

This applied a scoring approach, reflecting the preference of one team member. Evaluators believed that they were required to provide feedback on the program that would be based on an independent standard of performance justified by expert opinion.

The Niger evaluation

This followed a goal-oriented approach: the QA program was measured against its QA plan.

The Chile evaluation

This used a discovery approach. Evaluators sought to document lessons from the implementation of QA, looked for strengths in the way different districts had implemented QA, and engaged in dialogue with QA professionals to develop recommendations about the future of QA in the country.

Their method was influenced by appreciative evaluation, a method that was field-tested in that country (method described below).

All three evaluations included a retrospective evaluation with semi-structured interviews and record review [1–3]. However, the evaluators were surprised by how different their preconceptions were in each case and how this influenced both what they sought and found.

Upon reviewing these experiences, there was a compromise:

- (1) Avoid a formula-derived score for QA activities, because it would require far-reaching assumptions about one 'right way'—when clearly there were many—and also because it proved to be demoralizing for QA program participants, and
- (2) provide guidance for developing an approach and methods, and a menu of key aspects of QA activities, evaluation questions, and corresponding indicators to ensure some level of consistency between evaluations.

So, while there is flexibility within the guide, a clear set of key questions is recommended for different types of QA activities. The first two components of the guide were tested in Chile, and evaluators also tested two methods: a self-assessment tool and an appreciative evaluation method (presented below).

The evaluation guide

Part one: quality assurance evaluation implementation outline

This outline simplifies the complex task of managing an evaluation of QA activities and programs. It covers the basics of managing such an evaluation and provides suggested processes, sample forms, and checklists. The guide speaks of key challenges, including:

- (1) Addressing the needs of multiple customers (clients, stakeholders, and beneficiaries) makes it difficult to focus the evaluation
- (2) Managing the logistics is difficult and requires efficient collaboration with the hosts, including scheduling the in-country visit to accommodate the maximum possible number of sites, and accessing both staff and records
- (3) Selection methods must fit the evaluation purpose and also be practical (e.g. sample selection should take travel schedules and costs into consideration)
- (4) The outline presents the QA program map steps for recording QA structure and the existence and nature of QA activities [these steps are also presented in part two (matrix) in more detail; Appendix 1 presents the QA map steps]

The outline presents each major step in an evaluation, and provides ideas and suggestions. Its topics cover focusing the evaluation, managing a team-planning meeting, selecting a method, sampling, site selection, analyzing and presenting the results, and following up.

Part two: quality assurance evaluation matrix

The matrix presents a menu of key questions for each QA activity, suggested key indicators for each question, and a schedule, source of data, and proposed method to address the question (Table 1 presents a sample of the matrix). The purpose of the matrix is to help evaluators start planning for a specific evaluation. Once QA activities in a particular program are identified, the matrix provides a list of possible evaluation questions for consideration by evaluators, donors, and beneficiaries. As the evaluation questions are developed, the evaluator can tailor the matrix to the specific evaluation.

The matrix is organized by QA activity type. For each activity, the matrix follows a systems view of QA: the suggested questions and indicators aim to evaluate inputs, processes, outputs, and outcomes of QA activities and programs. The evaluators can select those questions that fit the needs of a particular evaluation (scope, relevance, client interest, and resources). The most helpful aspects of the matrix enable evaluators to select evaluation questions and to develop an evaluation plan; it also provides an outline for organizing the evaluation report. Improvements to the matrix resulting from the Chile tests involved refining and focusing several important questions related to the leadership of QA programs.

Part three: evaluation tools and methods

This section was developed as a reference for evaluators. It might be especially helpful if a program with few resources hires outside experts. It is not intended to replace basic texts, but rather to make the guide a complete reference tool. Methods included are: questionnaire-based interviewing, questionnaire-based self-assessment, record review, direct observation, cost evaluation of QA programs, and appreciative evaluation.

Findings from testing alternative methods

Once the lessons from Zambia and Niger were incorporated into the guide, we knew it had sufficient stability to permit research into ways to make it more useful and cost-effective. QAP tested two evaluation tools/approaches in Chile: self-assessment and appreciative evaluation.

Self-assessment was selected for testing because of two positive aspects: low cost and potentially positive effect on performance. Its reported shortfall is its tendency to render results with a positive bias. Appreciative evaluation was selected because of its potential to generate rich qualitative data, and its reported effects of increasing commitment to success and empowering those being evaluated.

In advance of the tests, QAP developed an evaluation instrument to assess these two tools against four dimensions: accuracy, validity, practicality, and effect on a client's ability to improve performance. Feedback was solicited from participants and evaluators immediately after the intervention.

Each method was tested in two of the 12 sites visited in Chile. Testing these methods was limited to the quality improvement activity only. The tests were not rigorous because of resource limitations, so results from the testing are

impressions of participants, donors, and evaluators, rather than scientific conclusions. As more evaluation tools are tested, new approaches can be documented to make the guide more complete. The testing methods and the results are summarized below.

Self-assessment

Description

Individual QA program participants answered questions in writing on a self-assessment form, with no prior consultation with each other or the evaluation team. Individuals rather than teams assessed their own performance.

Method

In advance of the evaluation visit, forms were distributed to test participants in two sites. Participants were asked to fill out both the form and an evaluation of the form. During the site visits to the two test sites, the evaluation team reviewed the results of the self-assessments and held a follow-up discussion with participants to discuss the accuracy and completeness of the data.

Results

Participants and evaluators alike felt that self-assessment was a useful method for soliciting information on certain types of questions—those where the participants had first-hand knowledge. For example, self-assessment findings in knowledge of QA, QA training, and performance after training were confirmed through the results of the group interviews. Evaluators concluded that self-assessment was not appropriate, however, for questions that required a wider knowledge of QA activities, such as their costs. Finally, participants recommended expanding the self-assessment form to enable them to provide their recommendations regarding the future of the QA program.

Appreciative evaluation

Description

This evaluation method contains a design 'bias' for seeking 'exceptional experiences' in an organization or system. Participants engaged in discussions that sought to identify the root causes of success in QA program implementation. This process required participants to interview each other based on a protocol, to report on each other's stories and recommendations, and to analyze the resulting data together. This method merges data collection and analysis by asking those who produce the data to participate in analyzing it and drawing implications from it.

Method

Background and instructions on the method were translated into Spanish and provided to the QA Director in Chile. She facilitated two appreciative evaluation sessions in two sites. In each case, a group of participants was asked to take part in the appreciative evaluation session, which lasted almost 3 hours. After the session, they were asked for feedback on that session, compared with more traditional group interviews.

Table 1 Quality assurance evaluation matrix (sample)

Evaluation question	When	Indicator	Source	Method
Management and supervision of QA What is the role of managers and supervisors in QA at every level of the health system? Explicit in QA program? Implicit in supervision?	Pre-visit and in-country		QA plan of QA manager	Document review or interview
Do supervisors have the competencies required to do QA supervision? Educational and training levels of supervisors? Competency evaluation?	Pre-visit and in-country	% supervisors with required competencies	Records Staff	Record review: comparison of recorded educational and training levels, and scope with required skill base Knowledge and skill testing through interviews or questionnaire
How often do they visit providers (for each level of the supervision system)?	Pre-visit and in-country verification	No. of visits per year (per facility) No. of visits per quarter (per facility)	Records Staff Supervisors	Record review Group interviews
Are they reviewing and supporting QA efforts, and if so, how?	In-country	% supervisors checking on or participating in QA activities	Staff Supervisors	Interviews with supervisors and providers
Do supervisors monitor compliance with standards?	In-country	% supervisory visits when compliance was checked	Supervisors Supervisees Record review	Individual or group interviews with supervisors Group interviews with supervisees OR self-assessment Tabulation of frequency of supervision visits
Do supervisors coach teams?	In-country	% of supervisors who describe coaching activities % of supervisees who report being coached	Supervisors Supervisees	Individual or group interviews with supervisors Group interviews with supervisees OR self-assessment
Do supervisors motivate teams?	In-country	% of supervisors who report that a supervisor contributed to their motivation	Supervisees	Individual or group interviews with supervisees

'Pre-visit' refers to work done by evaluators before visiting the country where the QA program is located; 'in-country' refers to the work done in that country.

The facilitator of the session and the evaluation team then provided their feedback to the R&D team. The results from these two sites were compared with results from the other eight sites where traditional group interviews were conducted instead.

Results

Both teams who participated were enthusiastic about using the appreciative evaluation methods. The results were similar to those of traditional group interviews in the other sites: everything discovered in a traditional group interview also came up in this session. The assumption that this would be a more empowering method was not validated, partly because the participants were already confident and enthusiastic (as compared with, for example, the Zambia program implementers who had been through a challenging re-organization and funding cuts). Appreciative evaluation has since been implemented in two evaluations with positive results. Information on appreciative evaluation will be presented in a panel discussion and a workshop at the American Evaluators' Association Annual Conference in November 2002.

Discussion

Overall

The greatest challenges in this R&D study were caused by the complexity of QA programs and the wide variation of country approaches. It was difficult to develop one consistent approach that would be applicable and useful in every situation. Evaluators of the three country QA programs agreed that the resulting menu of questions in the matrix was helpful and practical. Some QA staff, however, would have preferred a more clearly defined set of standards and thresholds for each QA activity.

If QAP continues in evaluation R&D, three changes are suggested: (1) explore tools and methods that help QA programs extract lessons learned from their experience; (2) test additional evaluation tools and methods; and (3) review and refine the matrix based on new frontiers explored by QAP (e.g. QA leadership and QA strategies for performance improvement).

Testing alternative methods

Test results for both self-assessment and appreciative evaluation produced positive impressions from participating local QA staff and evaluators about their usefulness. Findings from both methods were comparable to findings from group interviews. This is important for self-assessment, because it is a less expensive way of obtaining data. For appreciative evaluation, it demonstrated that the method did not 'veil problems', which is one of its critics' fears. More rigorous testing would be needed to understand the effects of these methods more accurately. They are both promising for different reasons: self-assessment is low cost, and appreciative evaluation can create better linkages between evaluation,

findings, and purposeful action (even though this was not demonstrated in the Chile evaluation).

Conclusions

- (1) Countries and programs differ significantly in the QA activities they emphasize, how their programs evolved, and the context in which QA is implemented. The QA evaluation guide recognizes and respects these differences, and offers a mapping tool and a menu of questions and indicators to facilitate the development of an evaluation method plan. Evaluators can adapt the indicators to meet their needs.
- (2) The evaluation guide follows a systems model and presents key questions and indicators for each QA activity. The guide follows the QA methodology that was developed and refined during QAP II (1997–2002).
- (3) Management and logistics are significant factors in the success of evaluations, especially those that are coordinated overseas. The evaluation guide provides guidance and ideas for how to plan and manage an evaluation more efficiently and effectively.
- (4) Self-assessment is a low-cost alternative for collecting some of the data in advance of the evaluation visit. Appreciative evaluation is an approach to concurrent data collection and analysis for engaging participants more fully in the evaluation. As more approaches are tested, they can be added to the tools and methods appendix.

Acknowledgements

This work was undertaken by the Quality Assurance Project with funding from USAID (contract No. HRN-C-00-96-90013) with the Center for Human Services, the non-profit affiliate of University Research Co., LLC.

References

1. Bouchet B, Fancisco M, Øvretveit J. The Zambia Quality Assurance Program: successes and challenges. *Int J Qual Health Care* 2002; **14** (suppl): 89–95.
2. Legros S, Tawfik Y, Abdallah H *et al.* Evaluation of the Quality Assurance Project and BASICS Joint Project in Niger. *Int J Qual Health Care* 2002; **14** (suppl): 97–104.
3. Legros S, Massoud R, Urroz O. The Chilean legacies in health care quality. *Int J Qual Health Care* 2002; **14** (suppl): 83–88.

Appendix I: mapping the existing QA system

Understanding and documenting where, how, and by whom QA is being implemented in a country is an important and

challenging task. QA activities and programs are complex and extensive; there is also wide variation among countries in the activities and roll-out of QA. Completing the documentation task or mapping of the QA system is the critical first step of the evaluation. To be successful at focusing the evaluation and interpreting results, one must first understand what is going on in QA, how it evolved, and the context in which it is operating. In mapping the QA system, the evaluator not only has to describe the system, but also has to analyze the significance and implications of the layout of the QA map. Here is a list of steps to follow:

- (1) Develop a comprehensive organizational chart of the health system showing all levels—central, provincial, and local—and identifying where QA occurs.
- (2) Identify where leadership for QA is located within that system (if it is) and how the other parts of the system interact with it: reporting relationships; formal and informal collaboration at the central level; links with the provincial, district and local levels; links and relationships with non-Ministry of Health (MOH) QA entities; and the informal power structure.
- (3) Document QA program goals, whether explicitly established in a QA plan or implicitly understood. If there is a QA plan, check and validate its goals to determine whether they are still applicable.
- (4) Review policies and procedures related to QA, staff assigned to it, the budget for QA services that have embraced QA, and the QA activities targeted by the country (both within the program being evaluated and in independent efforts).
- (5) Describe context characteristics including: (1) historical evolution of QA activities and programs that led to the current situation; (2) geographical factors (e.g. location, distance, infrastructure, communications) in the areas where QA is being implemented; (3) economic environment (e.g. gross national product and educational demographics); (4) policy environment such as policies about centralized and decentralized health and QA functions, resource allocation, regulations affecting health care, and the trade of health products and pharmaceuticals; and (5) cultural factors that may relate to patient and provider behavior and preferences.
- (6) The structure and capacity of the health system: address issues of centralization versus decentralization, supervision, training, procurement and logistics, and technology.
- (7) Organizational culture of the MOH and other key health institutions and their relationship to QA.

Accepted for publication 24 July 2002