

How Can Self-Assessment Improve the Quality of Healthcare?

by Sujata Bose, Elizabeth Oliveras, and Wendy Newcomer Edson

Executive Summary

Self-assessment could be particularly valuable in developing country healthcare settings where individual service providers often work without a supervisor or colleague to guide their performance. The QA Project has an interest in self-assessment and its impact on performance because of the intricate link between performance and quality.

This paper examines the issues relating to self-assessment, such as the different types of self-assessment, its uses, and its validity. It also reviews the literature (largely from developed countries) that informs our knowledge of self-assessment. The paper makes recommendations for future research and concludes that while much remains to be done to assure that self-assessment has the impact it promises, it may also be less costly and easier to implement than alternatives.

Introduction

Quality assessment is a vital component of maintaining and improving the quality of care provided by a health system, and assessing the performance of individual providers is an important part of system assessment. While clinic statistics can provide an overview of system performance, individual provider performance must rely on other information sources, such as assessments by supervisors, peers, independent external raters, or self. Supervisor assessment is the most traditional method, but is generally costly and sometimes impractical. In developing countries, supervisors often lack the resources they need to supervise effectively. For example, they may not have access to official vehicles, travel allowances, or service delivery guidelines. In addition, other clinical and administrative responsibilities

Recommended citation

Bose, S., E. Oliveras, and W.N. Edson. 2001. How can self-assessment improve the quality of healthcare? Operations Research Issue Paper 2(4). Published for the U.S. Agency for International Development (USAID) by the Quality Assurance (QA) Project, Bethesda, MD and JHPIEGO Corporation, Baltimore, MD.

Please note: This publication was previously released with the first author listed as “Sujata, B.,” rather than the correct author name provided above. The QA Project apologizes for this error.

The QA Project

The Quality Assurance Project serves countries eligible for USAID Missions and Bureaus, and agencies, and nongovernmental organizations that cooperate with USAID. The project team, which consists of prime contractor Center for Human Services (CHS), Joint Commission Resources, Inc., and John Hopkins University (JHU), provides comprehensive, leading-edge technical expertise in the research, design, management, and implementation of quality assurance programs in developing countries. CHS is a nonprofit affiliate of University Research Co, LLC, and provides technical assistance in the design, management, improvement, and monitoring of health systems and service delivery in over 30 countries.

JHPIEGO

The JHPIEGO Corporation, an affiliate of Johns Hopkins University, is a nonprofit corporation dedicated to improving the health of women and families throughout the world. JHPIEGO’s Training in Reproductive Health Project works to establish integrated (pre- and inservice) education and training systems to improve the performance of reproductive health professionals.

Financial support

The Quality Assurance Project is funded by the U.S. Agency for International Development (USAID), under Contract Number HRN-C-00-96-90013. Financial support for JHPIEGO was provided by the Office of Population, Center for Population, Health and Nutrition/Global Programs, Field Support and Research Bureau/CMT Division, USAID, under the terms of Award No. HRN-A-00-98-00041-00. The opinions expressed herein are those of the QA Project and JHPIEGO and do not necessarily reflect the views of USAID.

Continued on back outside cover

C O N T E N T S

- Introduction 1
 - What is self-assessment? 3
 - Benefits from self-assessment and its relation to quality assurance 4
- Conceptual and theoretical background 4
- Uses of self-assessment 6
 - Formative versus summative uses 6
 - Common uses of self-assessment 7
 - How self-assessment is conducted 9
 - Other levels of self-assessment 9
- Review of research 10
 - Method for literature search 10
 - Limitations of studies reviewed 10
 - Self-assessment for performance improvement 10
 - Development of self-assessment skills 11
 - Self-assessment for identifying learning needs 12
 - Validity 12
 - Outcomes of self-assessment 15
- Use of self-assessment in developing countries 15
 - Individual self-assessment 15
 - Group self-assessment 16
- Research recommendations 17
 - Which self-assessment instruments work best? 17
 - Which behaviors are more likely to be affected by self-assessment? 17
 - What is the effect of self-assessment on performance? 17
 - Can self-assessment work without supervision? 18
 - What is the role of feedback? 18
 - How should costs of self-assessment be measured? 18
 - Is a self-assessment program sustainable? 18
- Conclusion 18
- References 24



often take precedence over supervision, leaving little time for the observation of providers. Peer review is an interesting alternative that is receiving attention, but it is labor intensive and can create problems between staff members. Independent raters are widely used for research, but are costly. Self-assessment is a low-cost approach to monitoring the quality of care, and thus potentially important in developing countries, but questions have been raised about its validity.

Self-assessment is not new to healthcare. It is used regularly by clinicians to ensure that they act within their areas of expertise and to help them remain up-to-date with advances in medicine (Gordon 1992; Woolliscroft et al. 1993). It has been suggested that self-assessment is a prerequisite for maintaining professional competence (Das et al. 1998; Stuart et al. 1980). The literature cites an increasing need for self-assessment in healthcare settings as health systems change (Rooke 1998). Based on interviews and focus groups with persons who engaged in self-assessment, Marienau (1999) identified four types of benefits associated with self-assessment: learning from experience, functioning more effectively, strengthening commitment to competent performance, and fostering self-agency and authority.

There is an interest in formal self-assessment and its potential for improving healthcare services. In developing countries, significant numbers of healthcare workers function without supervision, particularly lone physicians and nurses at rural health posts and midwives and traditional birth attendants—many of whom work as solo agents. Because regular supervision is not possible in these dispersed systems, a mechanism by which workers could assess themselves between supervisory visits could prove an effective means of improving the quality of care. Giving providers a direct role in their own assessment could enhance the effectiveness of supervisory visits because the provider has already considered his or her performance and assessed its strengths and weaknesses. Self-assessment also has the potential to reinforce medical standards and to increase worker accountability.

The focus of this paper is individual self-assessment rather than assessment at the organizational or group level. The paper does, however, include group self-assessment because it has been used more often in health programs in the developing world.

What is self-assessment?

In 1977, Albert Bandura published a theory of the self-assessment process that includes self-observation, self-judgment, and self-evaluative reaction (Levine 1980). In other words, self-assessment involves observation of behavior,

evaluation of that behavior, and a reaction to the evaluation—more than simply measuring one’s own performance, it also includes an interpretation of that performance.

The total quality management (TQM) literature contains numerous definitions of self-assessment, but most refer to organizational self-assessment. The following definition of self-assessment from the European Foundation of Quality Management is typical of the TQM literature: “A comprehensive, systematic and regular review of an organization’s activities and results referenced against a model of business excellence” (Jackson 1998). The TQM definitions differ from those of individual self-assessment in two important ways: TQM uses a model or standard and culminates in planned improvement actions (Jackson 1998; Pitt 1999; Jackson 1999). In terms of healthcare, the use of a model for performance as an integral part of the self-assessment process is appealing and could easily be incorporated into individual self-assessments with practice guidelines fulfilling the role of a performance model. Action plans are often a part of the individual self-assessment process, too, though they are generally not included in definitions.

Abbreviations

AIHA	American International Health Alliance
ASHP	American Society of Health-Systems Pharmacists
AVSC	EngenderHealth (current name)
BARS	Behaviorally Anchored Rating Scale
COPE	Client-Oriented, Provider Efficient
CPI	Comparative performance information
GRS	Global Rating Scale
JHPIEGO	JHPIEGO Corporation, International Education and Training in Reproductive Health
MCAT	Medical College Assessment Test
MPQ	Management Practices Questionnaire
PBL	Problem-based learning
QA	Quality assurance
SESAP	Surgical Education Self-Assessment Program
TQM	Total quality management
URC	University Research Co., LLC
USAID	U.S. Agency for International Development

For the purposes of this paper, the following definition of self-assessment is used to focus on the individual in developing country settings:

Self-assessment is the ability of a health worker to reflect on his or her own performance strengths and weaknesses in order to identify learning needs, conduct a review of his or her performance, and reinforce new skills or behaviors in order to improve performance.

Benefits from self-assessment and its relation to quality assurance

Evidence shows that self-assessment by health workers may contribute to the goals of quality assurance in many ways. It

- Is a low-cost and sometimes valid method for quality assessment
- Can influence the health worker's behavior so as to increase compliance with standards (Adamow 1982; Love and Hughes 1994)
- Serves as an aid in professional development—clarifying areas for improvement, enhancing self-esteem, and developing self-awareness (Best et al. 1990)
- Gives participants greater ownership over the evaluation process
- Can improve communication between supervisors and subordinates (Harris and Schaubroeck 1988)
- May help to identify the transferable skills of workers (Mayall and Maze 1985)

All of these benefits are important in the healthcare setting. In less developed countries where resources are very limited and workers often must work on their own, the relative ease in executing self-assessment and its low cost make self-assessment especially appropriate. In addition, where supervisory systems are very hierarchical, increasing employees' participation in their own development is likely to improve the quality of their work overall. The increased participation of workers is in keeping with the paradigm shift in many international donor agencies that places increasing emphasis on health reforms and sustainable performance through a bottom-up approach.

Other potential benefits of self-assessment for quality assurance have not been well studied but are being pursued. The use of practice guidelines as a tool for self-assessment is one example. Including practice guidelines as part of the self-assessment package helps to make the guidelines available to providers and reinforces the standards. As medicine

moves toward an evidence-based model, the importance of adherence to guidelines is likely to increase. Some organizations in the U.S. already include practice guidelines as part of their self-assessment tools. For example, the American Academy of Pediatrics markets its self-assessment program as a tool for “evaluating the quality of your practice and improving your office's performance” (AAP 2000). Such combinations may be particularly valuable in developing countries as they can provide an additional means of making practice guidelines available to all healthcare workers.

The self-learning aspect of self-assessment is particularly important in health because clinicians must stay abreast of advances in patient care. The Surgical Education Self-Assessment Program (SESAP) of the American College of Surgeons was designed with this in mind. It was believed that surgeons would educate themselves by reviewing the questions they answered incorrectly (Rosato 1972). The SESAP continues to be used in this way both as a study guide for in-service exams and as a means of earning continuing medical education credit (Lux 2000).

Self-assessment can be an inexpensive, easy-to-use tool for determining whether providers are performing according to standards and for correcting those behaviors that are not consistent with standards. In this capacity, self-assessment may be a useful supplement to supervisory systems that lack regular direct supervision. There are, however, a number of limitations to using self-assessment in this manner: without careful attention to its appropriate use, self-assessment is unlikely to meet its intended goals.

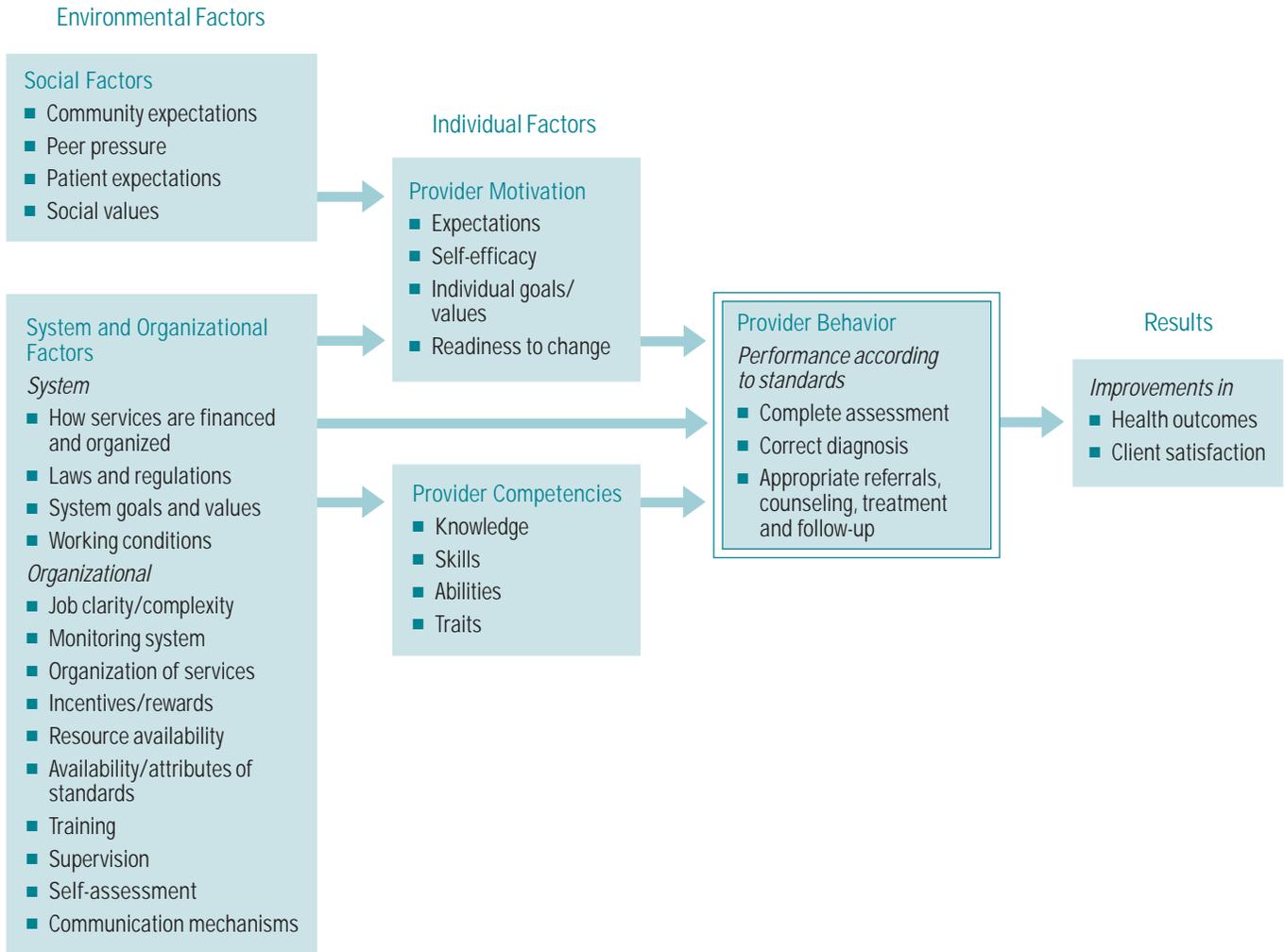
Conceptual and theoretical background

The QA Project has developed a conceptual framework for the determinants of health worker compliance with standards that takes into account individual, organizational, and social factors that interact to influence provider behavior and their success in performing according to standards and, ultimately, health outcomes (Figure 1).

In this framework, self-assessment is considered as an organizational factor that influences provider competence and motivation. Self-assessment can affect provider competence by reinforcing knowledge and improving skills. It can influence provider motivation by strengthening self-efficacy and readiness for change. Readers interested in a fuller discussion of measuring health worker competency and performance according to standards are referred to Kak et al. (2001) and Marquez (in press), respectively.

Interest in self-assessment in the U.S. began in the 1970s. Disciplines such as education, health, organizational/

Figure 1 ■ Determinants of Healthcare Provider Performance According to Standards



industrial psychology, and management used self-assessment in large corporations, banks, factories, colleges, and elementary schools. In the management field, it developed in part out of the movements toward participative management (Thornton 1968) and TQM, and in the health field from the need for continued self-learning throughout one's career (Woolliscroft et al. 1993).

Although some theoretical work related to self-assessment exists, multiple authors have suggested that more sound theoretical grounding would be useful (Heneman 1980; Gordon 1992; Korsgaard 1996). For example, there is a need for further work on theoretical constructs and the linkages between theory and application.

One body of theoretical work addresses factors that motivate self-assessment behavior. Table 1 summarizes three

theories that link self-assessment to self-esteem. Each of the theories reaches a different conclusion about the bias associated with self-assessment. These sources of bias in self-assessment are key to understanding the limitations of its use.

Trope and Pomerantz (1998) suggest that there is a conflict between: (a) the desire for self-assessment accuracy as proposed by self-efficacy, and (b) defensive motives related to self-esteem as proposed by the self-enhancement and consistency theories. Self-enhancement theory predicts that self-assessors will distort evaluative information processing to support their desire to see themselves favorably. Consistency theory predicts that individuals are more inclined to remember and recall behaviors that are consistent with their self-esteem.

Table 1 ■ Theories That Address Self-Assessment

Theory	Description	Bias
Self-efficacy theory (Bandura 1977)	Knowledge of personal competence is gained largely from observing and evaluating one's own behavior and its effects.	People are motivated to assess themselves accurately. (No bias)
Self-enhancement theory (Greenwald 1980)	People desire to see themselves favorably and as competent.	Low self-esteem self-raters will evaluate themselves more favorably than high self-esteem self-raters, given identical performance. (Leniency bias)
Consistency theory (Korman 1970)	People perceive their behavior in consistency with their self-esteem in order to maintain a consistent self-image, which is more psychologically pleasant.	High self-esteem self-raters will evaluate themselves more favorably than low self-esteem self-raters, even if their behavior is identical. (Leniency bias)

Wells and Sweeny (1986) consider whether self-esteem is stable and what this means for self-assessment. Kernis et al. (1996) claim there is “growing concern that individuals’ self-appraisals have both a stable and unstable component.” While it is not clear whether the stability of self-esteem is a function of the dimensions people use to judge themselves (e.g., competence, athleticism) or a function of the type of judgment, it appears that stability is related to self-assessment.

Another body of theoretical work addresses how self-concepts and, therefore, self-assessments are formed (Table 2). Farh and Dobbins (1989a) suggest that social comparison is used in self-assessment because the criteria used to judge people are rarely absolute. Symbolic interactionism holds that when people assess themselves, they often try to do so from the perspective of others.

Uses of self-assessment

Formative versus summative uses

Self-assessment is used for both formative and summative purposes. Best et al. (1990) describe the difference: “A formative evaluation is an internal evaluation which serves to improve the product being developed [while] a summative evaluation is an independent assessment of an outcome, which is a judgment of worth of the final product.” Although authors agree that self-assessment can be useful for formative purposes, they disagree about its role for summative purposes. The gist of the argument against using self-assessment for summative purposes is that asking people to assess themselves when rewards and punishments are involved is unfair and ultimately inaccurate. The price of personal objectivity and honesty may be too high (Fuhrmann and Weissburg 1978, cited in Best et al. 1990; Arthur 1995).

Table 2 ■ Theories That Address Self-Concept

Theory	Description	Testing
Social comparison theory (Festinger 1954)	There are two standards for abilities: physical and social. When physical criteria do not exist, individuals compare themselves with others to determine their ability levels.	Klein (1997) studied whether social comparison information is used when more objective information is available, concluding that both objective and social measures are used for self-assessment if both are available.
Symbolic interactionism (derived mainly from George Mead's 1934 work <i>Mind, Self and Society</i>)	Individuals develop self-notions by placing themselves in others' positions to better understand their perspectives, thus learning to view themselves more objectively.	This theory predicts that, because conceptions of self derive from the same processes as perceptions of others, there should be congruency between self-perception and perceptions by others (John and Robins 1994).

Those who support self-assessment for summative purposes point to evidence that students can self-assess accurately (Henbest and Fehrsen 1985), and argue that summative self-assessment can improve the job performance appraisal process (Thornton 1980; Makiney and Levy 1998; Korsgaard 1996). For example, Thornton (1980) concludes that self-assessment by workers reduces defensiveness, highlights areas of disagreement with supervisors, and leads to improved job performance.

Formative uses of self-assessment focus on individual learning, particularly to reinforce behavior change. Ensuring the continued quality of services requires ongoing monitoring and evaluation. Providers who monitor themselves acquire greater buy-in to the system as a whole. Self-assessment may also identify issues that would not be raised by outside observers who can see only a limited amount of service provision.

Common uses of self-assessment

Table 3 summarizes the four major uses of self-assessment: identifying learning needs, improving performance, appraising performance, and reinforcing skills. In reality, these are not distinct categories—performance appraisal feeds into performance improvement, and identifying learning needs can lead to skill reinforcement. They are, however, useful distinctions for organizing the empirical evidence on self-assessment and identifying how self-assessment can improve healthcare systems.

Other terms for self-assessment

Self-evaluation and **auto-evaluation** (much less common) are used synonymously with self-assessment.

Self-reflection (more common in the nursing literature) is a process that focuses on the motivations and feelings surrounding actions rather than the assessment of performance (Rooke 1998). **Reflection** is “a process of reviewing an experience of practice in order to describe, analyze, evaluate, and so inform learning about practice (Reid 1993, cited in Rooke 1998).” Brew (1999) states, “All self-assessment involves reflection, but not all reflection leads to self-assessment.”

Self-measurement or **self-testing** is the use of an objective, external standard (e.g., a scale, test questions) to measure performance defined as outcomes. Self-measurement does not include the interpretation of performance, a key component of self-assessment, but can include self-measurement as a tool for assessing performance.

Self-learning is a different concept than self-assessment, but self-assessment can contribute to self-learning (e.g., by identifying weaknesses).

Table 3 ■ Major Uses of Self-Assessment

Use	Description	Example
Identifying learning needs	Used to help students and practitioners delineate areas where they feel they need to improve their understanding or require further training	Self-assessment programs sponsored by several professional medical organizations aim to help members identify their strengths and weaknesses, and compare their knowledge and competence to that of their peers.
Improving performance	Used to change a person’s behavior by changing his/her perception of the behavior	One physician developed a self-assessment process for the members of the group practice he managed as a means to help the physicians change their own behavior (Flood 1998).
Appraising performance: both ongoing monitoring of performance and evaluation of performance	Used to review performance as an alternative to external monitoring and supervision at the work site and in training programs	Some US companies use self-assessment and goal setting for employee performance review. Some medical and nursing education programs have their students self-assess their performance in courses.
Reinforcing cognitive abilities and skills after training	Used to sustain changes in the performance of providers after training	Providers in Indonesia who engaged in self-assessment and peer review after a training on client-centered family planning counseling maintained improved performance after the course significantly better than the controls who received training only (Kim et al. 2000a, b).

Identifying learning needs: Self-assessment helps students and practitioners delineate areas where they feel they need to improve their understanding or require further training. This information can guide an individual to pursue more information or training, or it can be used by organizations to plan education or training programs. For example, self-assessments have been used by health organizations since the 1970s to identify areas where their members need upgrading (Rosato 1972). Several medical associations, including the American Psychiatric Association, the American Academy of Pediatrics, and the American College of Obstetricians and Gynecologists, have instituted similar programs (Parboosingh 1998). These programs let individual practitioners know where they stand in comparison to their colleagues and identify topics where they may need refreshing. It also allows the professional association to determine overall areas of knowledge in which their members may need further training.



Increasingly, self-assessment is being viewed as an alternative to external monitoring and supervision, and as such, it falls within the context of performance appraisal.

Evidence from within and outside the health field indicates that self-assessment is a useful method of self-learning. In an innovative program with eighth-grade science students, Ballenger (1974) negotiated a program of mutual self-assessment and performance improvement for students with a history of underachievement. In the first term, mean tested performance in the science course fell below performance in other courses, but in the second term, science performance climbed above that of other courses and above previous science performance. Improvement was attributed to more time spent on tasks and more material covered by the self-assessing class. Here, self-assessment provided a valuable impetus toward learning; it yielded measurably improved academic results, using objective testing instruments to gauge academic performance.

Individuals may be threatened by self-assessment, so establishing an open and trusting environment is essential to its success. Kennell et al. (1973) report that medical school teachers who took part in a self-assessment program to improve awareness of their own and others' teaching styles felt vulnerable to course leaders' criticisms, because they were not convinced of their neutrality. Some learners wrote scathing letters or refused to participate.

Improving performance: Some studies report that self-assessment can be effective in causing desirable behavior change. One physician developed a successful self-assessment process for the members of the group practice he managed as a means to track and measure changes in physician behavior and the effects they have on the practice of medicine (Flood 1998). Ballenger's 1974 study of eighth-grade science students is another example. Several assumptions underlie the belief in self-assessment as a means to behavior change. Best et al. (1990) hold that learning that influences behavior is self-directed and self-discovered. Parboosingh (1998) distinguished between formal self-assessment (written tests of knowledge) and self-reflection, and concluded that self-reflection was more likely to result in a commitment to behavior change.

Appraising performance: A number of studies look at the effect of self-assessment on the appraisal process, including formal performance reviews. Involving employees in their performance reviews is increasing as part of a more participative approach to management (Thornton 1980). Klimoski and Inks (1990) found that supervisors' ratings of a subordinate's performance were swayed by subordinates' self-assessments, but anticipation of face-to-face feedback did not lead to inflated appraisals. Makiney and Levy (1998) found that supervisors were more likely to incorporate information that conflicted with their own evaluation if that evaluation came from a peer than if it came from the employee himself. Discrepant self-assessment information was not incorporated into the supervisor's final judgment.

Self-assessment has been adapted to serve a monitoring or evaluation function in both work sites and training programs (e.g., medical schools). Increasingly, self-assessment is being viewed as an alternative to external monitoring and supervision, and as such, it falls within the context of performance appraisal. Much of the literature on self-assessment for monitoring and evaluation focuses on the validity and reliability of self-assessment. Even though self-raters may systematically under- or overestimate their own performance, such errors may not result in biased estimates of *change* in performance, such as when used in an ongoing performance-monitoring process.

Self-selection generally refers to the ability of individuals to determine whether they are suited to a particular job. One study in this area used self-assessment with Peace Corps trainees to help them determine if they should continue or withdraw from training (Katz 1970, cited in Gordon 1992). The field performance of those who completed the training was found to be superior to Peace Corps volunteers who had not used self-selection.

Reinforcing cognitive abilities and skills after training:

Self-assessment has been shown to sustain new behaviors learned during training. In a study by the QA Project, Kim et al. (2000a, b) introduced self-assessment and peer review to family planning counselors in Indonesia following a one-week training on counseling skills. After 16 weeks, counseling performance was significantly higher in the groups using self-assessment and peer review than in the control group.

Use of self-assessment to improve cognitive abilities is not common; instead, efforts are geared toward identifying learning needs. Where this is effective, self-assessment can reinforce cognitive capabilities and specific areas of knowledge. In Ballenger's (1974) successful intervention with science students, the students selected the areas in which they wanted to concentrate.

How self-assessment is conducted

The most common way to conduct self-assessment is to have people fill out a brief questionnaire or checklist on paper; other methods are a detailed questionnaire, journal and diary entries, or review of a videotaped performance. Ratings against criteria of behavior are the most common form of written assessment, although full, written responses are sometimes used. Occasionally, particularly in an educational context, self-assessment is an oral exercise, carried out in front of one's peers or fellow students. Journals and diary entries can be used to reflect on and learn from or about one's performance. The use of audiotape and videotape to enhance the self-assessment/self-learning process has been tried, and computer-based self-assessment programs developed. Taylor (1998) discusses the use of computer-based self-assessment for math students; the analysis of the software program showed it to be acceptable to students and easy to use.

Self-assessment is usually carried out in conjunction with support from or an evaluation by a supervisor. Evaluations by supervisors or even peers have often been used to validate self-assessments, particularly where self-assessment was carried out for monitoring or evaluation purposes. Even where self-assessment is used for learning purposes, it is common to support it with supervisor feedback and validation, which gives employees and/or students a more objective basis for developing future performance goals. In many cases, the employees or students spend time discussing and reconciling their assessment with that of an observer as part of the learning process.

Self-assessments that do not include supervisor support generally provide some other form of external support. Assessments by medical organizations in the U.S. are geared

to identify strengths and weaknesses, and are sometimes used for recertification of clinicians. They often include a number of elements beyond the self-assessment tool, such as a literature review, and report back results (Parboosingh 1998). Providing clinicians with information on their results and their peers' allows them to compare their performance with that of others. Literature reviews are included as a way for clinicians to refresh their knowledge and study those areas where knowledge was lacking. In some cases, clinical guidelines or standards are also included as a means for updating and refreshing knowledge.

The literature yielded only one account of self-assessment attempted without any support (Maguire 1990). It concluded that having students watch a videotape of their performance did not enhance their learning communication skills.

Other levels of self-assessment

While individual self-assessment is the focus of this paper, other types of self-assessment do exist and are relevant to quality assurance in healthcare. Self-assessment can be performed by teams, facilities, and organizations to assess their performance. In fact, this type of self-assessment has been used by numerous organizations throughout the world. Group self-assessments in the health field have been conducted for many purposes, including determining how to improve organizational performance (Jackson 1999), stimulate positive learning and improvement (ASHP 2000), and evaluate programs (AIHA 2000). In international health, group self-assessment is more prevalent than individual self-assessment.

The objectives of group self-assessment have generally arisen from TQM initiatives and are focused on continuous quality improvement. They are, however, quite similar to those of individual self-assessment. Attention is given to identifying strengths and areas for improvement and to develop self-assessment skills (Pitt 1999). Group self-assessment differs fundamentally from individual self-assessment in that the former's unit of analysis is larger: a clinic, a hospital, or a unit in a hospital. In many cases, efforts are made to include staff at all levels of the organization in order to give a more complete representation of the services being provided. In addition, information is sometimes collected from other sources (e.g., output data, client interviews) to complement that provided by staff.

Group self-assessments tend to follow a slightly different process, though they often incorporate tools similar to those used for individual self-assessment. Common additions for group self-assessments are discussions of the individual assessments of the group and organization and

record reviews. The COPE (Client-Oriented, Provider Efficient) tool, developed by AVSC International,¹ includes client interviews, a client-flow analysis, and development of a plan of action (Lynam et al. 1993). Data from these sources are considered in the meetings of all participating staff. Such meetings are an important part of the group self-assessment process as they allow everyone involved to have input not only in terms of data, but also in terms of identifying problems and solutions. “A broad multi-disciplinary discussion is critical to reach conclusions and to avoid similar problems in the future” (Myhre 1998).

Review of research

Method for literature search

The studies included in this paper were found through a literature search using Medline, ERIC, PsychLit, CINAHL, and Popline to search for the terms *self-assessment* and *self-evaluation*. The search focused on the years 1990 to 2000 and on healthcare literature. Additional articles were identified and retrieved by reviewing references. Several excellent reviews and meta-analyses were particularly useful in identifying references (Arthur 1995; Falchikov and Boud 1989; Gordon 1991; Harris and Schaubroeck 1988; Mabe and West 1982). Because of the relatively small number of articles on any given topic (excepting validity and accuracy of self-assessment), strict criteria for inclusion were not used.

Limitations of studies reviewed

Most of the research on self-assessment has been done in the U.S. and Europe. While the goal of this paper is to help those working in international health draw conclusions about the applicability of self-assessment in developing countries, there is little research in this setting. What does exist is incorporated as appropriate.

A drawback of the research on self-assessment is its conduct in experimental settings. Many of the studies rely on laboratory settings or fictional scenarios to investigate elements of self-assessment. Such studies may not give an accurate picture of the actual use of self-assessment. Recent literature on the reliability of self-assessment includes an important critique of the statistical techniques that were previously used to assess it (Atwater and Yammarino 1997; Johnson and Ferstl 1999).

Most of the empirical evidence on self-assessment addresses issues of effectiveness and validity. Other studies address related topics, including factors that affect its implementation. Within health, much of the literature focuses on self-assessment in educational settings; relatively few studies consider the use of self-assessment in clinical or field settings. Many of the studies of self-assessment come from outside the health field, so while this paper focuses on the evidence related to self-assessment in healthcare settings, examples from other fields are included to provide a more complete picture.

Self-assessment for performance improvement

To date, relatively little research has been conducted on the use of self-assessment for performance improvement. The studies that are most relevant are those that look at whether self-assessment is an effective tool for creating behavior change. If not, it will not be useful as a means for improving clinicians' performance.

The few studies that look at the use of self-assessment for changing particular behaviors suggest that it is effective for some, but not all, behaviors. Sideris et al. (1990) looked at the effectiveness of self-evaluation for improving the interviewing skills of doctors. Fifteen doctors used a self-evaluation checklist to identify their errors while listening to tapes of patient interviews they had conducted. Overall, they showed improvement in some communication skills but not in written history skills. The authors conclude that history-writing skills did not improve because the self-evaluation was not oriented toward changing this behavior. A study by Mason et al. (1988) looked at the use of a self-education video by medical students for improving interview skills. One of the study groups used self-evaluation while the other three did not, and the self-evaluation video had no discernable effect.

Overall, these examples suggest that self-assessment can be linked to improvement in performance despite its low validity (discussed below). It may be that self-assessment provides an effective means for individuals to reflect on their performance and devise ways to improve it.

Another behavior change that has been shown in a number of studies is an increase in communication between the learner/employee and the trainer/supervisor (Gordon 1992). An increase in positive communication can be important to the performance improvement process because it may make employees feel more in control of their performance.

¹ AVSC has changed its name to EngenderHealth since developing COPE. Please see box on COPE on page 16 for more information on this methodology.



Development of self-assessment skills

The ability to self-assess was once assumed to be inherent, particularly to health professionals. More recently, a number of authors, in and outside the field of health, have concluded that self-assessment skills must be learned (Best et al. 1990; Falchikov and Boud 1989). "Students need systematic practice in judgment of their own work and feedback on the ability to do so" (Brew 1999). A meta-analysis of self-assessment in higher education concluded that self-assessment skills had to be developed (Falchikov and Boud 1989). This analysis identifies areas where training may be most useful, particularly discussion, understanding, and explicitness of evaluation criteria.

Mabe and West's 1982 meta-analysis of self-assessment concludes that "self-evaluations become more accurate as subjects gain experience in evaluating their abilities." Gordon (1991) reached a similar conclusion in his review of the validity of self-assessment. Sciabassi and Woelfel (1984), on the other hand, looked at performance at different stages of an anesthesiology clerkship to see if self-assessment ability improved over time. They found no trend toward improved correlations between student and instructor ratings over time. Given that student and teacher assessment may map different dimensions, this finding is not necessarily an indication that self-assessment skill did not improve.

Hay (1995) conducted a study of the effectiveness of self-evaluation during a problem-based tutorial for the development of self-assessment skills. Thirty students used self-evaluation throughout a 14-week occupational therapy course. Although the correlations between student and tutor evaluations (completed using the same forms) increased over time, the author suggests that this resulted from a negotiation between the tutors and students, with the tutors changing their evaluations to more closely reflect those of the students, rather than from actual learning of self-assessment skills. Cochrane and Spears (1980) had dietitian students repeatedly self-assess their performance during a course, while also providing the students with periodic faculty evaluations of their performance. The differences between the student self-assessments and the faculty evaluations decreased during the course, although it is not clear whether this was due to increased learning by the students.

Assuming that self-assessment skills can be taught, some authors have attempted to determine the best way to do so. Gordon (1991) suggests that clear goals and explicit feedback regarding skill at self-assessment are vital to their

development. Students taking a course on fire science and management that included self-assessment recommended that the course provide more frequent feedback (Trevitt and Pettigrove 1995). A study of physical therapy students compared self-assessment with and without videotape review and found that students who reviewed the video gave more detailed criticism but were no more accurate (Palmer et al. 1985, cited in Gordon 1991). This suggests that self-observation does not necessarily improve self-assessment skills. A Canadian nursing program (Best et al. 1990) used a model of collaborative evaluation, whereby teachers in a nursing program effectively mentored students and helped them to become proficient self-evaluators by talking through problems and helping to set clear standards, criteria, and objectives. It is unclear whether this approach improved self-learning, but it is in keeping with the idea that practice improves self-assessment skill and that training and clarification of objectives make self-assessment more effective.

Using another approach, Adams et al. (1974) had peers of medical school teachers observe them teaching, and afterwards in group discussion the teachers assessed their own performance and the peers responded. This approach encountered more problems than private self-assessments by students, as many of the teachers were suspicious of the motives behind the exercise and were uncomfortable with the format.

Despite some evidence to the contrary, the evidence generally suggests that self-assessment skills can be taught and must be taught if self-assessment is expected of practicing clinicians. One limitation of the research in this area is the lack of studies that address whether self-assessment skills, once learned, are applied on the job.

Other theoretical literature on self-assessment focuses on self-evaluation as a complex skill that must be learned and practiced over time (Best et al. 1990). A 1978 interpretation by Zabarenko and Zabarenko (cited in Arthur 1995) developed this further, framing self-assessment as a skill that evolves within the context of specific professions. They state that skill in self-evaluation may be linked to the development of one's self-concept and how one views him- or herself as a professional. For example, for a nurse the ideal

professional nurse might be characterized by standards of practice, a specific knowledge base, and other behavioral and personal imperatives resulting in a sense of internal satisfaction and being valued by society. Arthur writes that, while the basics of self-assessment may be operative from the start, it is only when students have taken on the ego-ideal of the profession that they can closely scrutinize the “self” and restructure their goals accordingly.

Self-assessment for identifying learning needs

There is surprisingly little published research on the contribution of self-assessment to self-learning. In a review of 57 self-assessment studies, Falchikov and Boud (1989) found that relatively few had self-learning, or analogous aims, as a study objective. Only 15 percent were done to increase effectiveness of learning, to improve the learning experience, or to promote reflection on the part of student self-assessors. Four percent aimed to measure participants’ attitudes toward self-assessment and examine its effect on learning. (It was not possible to tell whether the 4 percent and 15 percent overlapped or were independent.) There is some documentation of how self-assessment contributes to the self-learning process. Katz (1970, cited in Gordon 1992) found that Peace Corps trainees eventually accepted responsibility to direct their own learning and make their own decisions. This suggests that self-assessment can lead to greater responsibility for the direction of learning.

Self-assessment can improve self-learning in several ways. One is by increasing communication between self-assessors and others involved in their work. For example, in a study with dietetics students, Cochran and Spears (1980) initiated a program of self-assessment in which specific behaviors were identified that were to be encouraged or improved through immediate feedback. Students and instructors independently assessed and then met to reconcile their differences of opinion. The process increased student initiation of communication and feedback sessions with faculty. Stackhouse and Furnham (1983) reported a similar increase in student initiation of communications and feedback sessions with faculty, using a program of self, clinical supervisor, and academic tutor rating and conferral for speech therapy students. Gordon (1992) says that benefits included improved morale, motivation, and communications among learners and faculty. Overall, there is a sense that the self-evaluation process enables participants to initiate more dialogue with supervisors and teachers. This may make it easier for both sides to delineate specific suggestions for improvement, thus aiding the self-learning process.

Validity

Do self-raters judge themselves accurately?

Much of the literature on self-assessment focuses on the question of validity: How well does the self-assessment agree with an external “objective” evaluation by others? The more objective standards against which studies commonly compare self-assessments are supervisor and peer evaluations. In academic settings, comparisons have also been made against grades and standardized tests and exams. Whether or not these standards are actually the “gold standard” against which self-assessments should be judged has been questioned, but new studies continue to look at this issue.

The results of studies looking at the degree of validity of self-assessment have been ambiguous. The most common finding is that self-assessments have low to moderate validity compared to external measures of capability and performance (Arnold et al. 1985; Sclabassi and Woelfel 1984; Arthur 1995; Reilly and Chao 1982; Harris and Schaubroeck 1988; Fincher et al. 1993; Frye et al. 1991; Lichtenstein and Fischhoff 1977; Hunter and Hunter 1984; Risucci et al. 1989; Gordon 1991). However, some studies show higher validity (Kaiser and Bauer 1995; Pym and Auld 1965; Williams and Seiler 1973). These results pertain to both the medical training literature and the nonmedical literature; all studies found in the medical literature related to training.

In the medical literature, studies report low validity when the external assessment was by the supervisor or trainer, or an examination score. Sclabassi and Woelfel (1984) found no significant correlation between student and instructor evaluations using a sample of 130 medical students. Arnold et al. (1985) found low correlations between student self-assessments and grades on both medical school exams and National Board of Medical Examiners exams. In a review of self-assessment in the healthcare literature, Arthur (1995) concluded that self-evaluation was associated with 15 percent or less of the variance in all (peer or faculty) measures of student performance.

Numerous authors have studied whether individuals tend to underestimate or overestimate their own abilities and performance in relation to external raters. The general consensus is that more tend to overestimate (Thornton 1980; Harris and Schaubroeck 1988) and this is the reason for the low correlation between self-assessment and assessments by others. Falchikov and Boud’s (1989) meta-analysis of self-assessment in higher education showed that students overrated themselves compared to their teachers. In a meta-analysis of ratings by peers, supervisors, and self, Harris and Schaubroeck (1988) found higher correlations between peer and supervisor ratings ($r = .62$) than between

self and peer ($r = .36$) or between self and supervisor ($r = .35$).

However, some individuals underestimate their abilities (John and Robbins 1994). For example, medical students appear to be more likely to underestimate their ability compared to their teachers and other supervisors (Sclabassi and Woelfel 1984; Fincher et al. 1993; Frye et al. 1991; Morton and Macbeth 1977; Arnold et al. 1985). Hinsz and Matz (1997) found that 20 percent of their subjects rated themselves below average, a higher rate than reported by other studies, possibly suggesting that underrating is more common than suspected. Farh and Dobbins (1989b) found that high self-esteem self-raters exhibited more leniency bias than low self-esteem self-raters. Differing information environments and emphasis on role requirements may influence the duration of the bias (Shore et al. 1998).

Several review articles and meta-analyses have come to the same conclusions. In a review of 18 studies of self-assessment in health profession training, Gordon (1991) concluded that self-assessments are most often tied to self-concept of “global attribution of ability” rather than to actual performance. Falchikov and Boud’s (1989) meta-analysis of self-assessment in higher education showed that students overrated themselves compared to their teachers. Mabe and West (1982) concluded that the literature on self-assessment does not give a clear answer about validity with respect to specific skills or performance categories.

Self-enhancement has been the most studied source of bias (Shore et al. 1998) and has been supported by a number of studies (Sedikides 1993). John and Robbins (1994) tried to determine whether all individuals self-enhance and found that while some do, others do the opposite. They did reach the common conclusion that people can assess others more accurately than they can assess themselves. Other studies provide support for self-consistency theory.

Does validity matter?

A number of authors claim that validity may be less important than the literature suggests. Nowack (1992) proposes that there may be advantages to individuals serving as judges of themselves. He refers to Shrauger and Osberg (1981) who suggested that self-assessment may be a better predictor of some behaviors than reports by others. Heneman (1980) says, “[T]here is apparently little theoretical concern with why one might expect self-assessments to significantly correlate with the criterion in the first place...” Even when reliability is low, ratings may still accurately reflect the rank ordering of performance dimensions within ratees and the rank ordering of overall performance across ratees (Cheung 1999). This points to the usefulness of self-assessment even if validity is low.

Some authors have looked at causes of discrepant ratings, arguing that it is not just the levels of ratings that matter: the reason(s) for the difference must be taken into consideration if inferences are to be made about self–other differences. Cheung (1999) reviewed various explanations for discrepant ratings, including conceptual disagreement (that raters use different frames of reference for performance evaluation) and psychometric (that raters respond to a scale in different ways). He also describes methods for testing to examine these disagreements. “Other equally plausible, but less studied explanations for discrepant self-rating have been proposed, including varying information environments, and differential weighting of available information.”

Martin (1998) proposes three general reasons for the low correlation between self-ratings and other ratings.

- Individuals evaluate themselves on different dimensions of performance than do experts (Farh and Dobbins 1989b; Hughes et al. 1997; Johnson and Ferstl 1999)
- Different individuals may perceive the same self-assessment score differently (e.g., seven out of 10 may be seen by some as a superior rating and by others as average)
- Because self-assessors have not seen the full range of competence, they may use different benchmarks for quality than experts (Farh and Dobbins 1989b)

Kruger and Dunning (1999) argue that “the skills that engender competence in a particular domain are often the very same skills necessary to evaluate competence in that domain.” A series of small experimental studies supported their hypothesis that incompetence not only causes poor performance but also leads to the inability to recognize it.

Several methodological weaknesses have been identified in studies of validity (Falchikov and Boud 1989; Mabe and West 1982; Atwater and Yammarino 1997; Johnson and Ferstl 1999). Mabe and West (1982) suggest that inefficient handling of measure reliability and restriction of range have been factors. Falchikov and Boud (1989) also point to differing interpretations of “agreement” as a cause for findings of low validity.

Atwater and Yammarino (1997) argue that the type and degree of agreement between self-ratings and other ratings affect outcome, such as job attitudes, self-diagnosis of training needs, and goal setting and attainment. As a corrective measure, they propose greater use of feedback in light of research showing it increases agreement between self- and other assessments.

Factors that influence the validity of self-assessment

Gordon (1997) analyzed the factors that influence the validity of self-assessment, concluding that “the validity of self-assessments appears to improve when two conditions are met: (a) learners are able to systematically gather and interpret data on their own performances, and (b) faculty and trainees have the chance to compare and reconcile their independent assessments of the learners’ performances.” More specific factors that have been found to affect the accuracy of self-assessment include:

- Explicit criteria or standards for the procedures to be evaluated
- Experience in self-evaluation
- Expectation of self-assessment validation (Gordon 1991; Mabe and West 1982; Myhre 1998)

A few studies have tested specific factors posited to improve the accuracy of self-assessment. Farh and Dobbins (1989a), for example, tested the ability of social comparison information to improve the validity of self-assessment. They had 163 undergraduate students complete a series of editing tests. The students were broken into groups, and those in the social comparison group were allowed to review the work of other “editors” before they were asked to rate their own performance. The study found greater correlations between self-ratings and objective performance indicators when comparative performance information (CPI) was available to ratees. The authors conclude that “[CPI] should lead to greater feedback acceptance and positive behavior change.”

Stuart et al. (1980) compared the ratings of residents and faculty and found that clear and detailed criteria increased accuracy.

Other factors that have been suggested to affect self-assessment validity are:

- Self-evaluation instructions using social comparison terminology
- Instructions of anonymity
- Intentional, positive incentives to improve validity
- Knowledge of local or individual results
- Employee’s length of service
- Employee’s familiarity with subject matter (Love and Hughes 1994)

Stahl (1998), in his discussion of self-assessment as part of the process of quality assurance for continuing training,

discussed obstacles to honest self-evaluation. In particular, he focused on the need for organizations implementing self-assessment to break the link between lack of skills and negative judgment. In order for employees to be honest in acknowledging their weaknesses, they must know that they will not be penalized for having done so.

Myhre (1998) goes beyond the accuracy of the self-assessment and considers what will make the process of self-assessment most effective in terms of effecting change. He suggests that it is key to have systems that permit corrective actions when deficiencies are observed.

All told, the factors that seem to be the most influential are explicit, well-understood criteria; practice with self-assessment; anonymity; and feedback/validation. Accuracy may vary by an individual’s characteristics (e.g., length of time on the job).

Interaction of feedback and self-assessment

Johnson and Ferstl (1999) studied the effect of subordinate feedback on the performance of 2,171 managers. When a manager’s initial self-assessment exceeded subsequent ratings of the manager by subordinates, the manager’s performance tended to improve after the feedback of the subordinates’ rating. However, when a manager’s initial self-assessment was lower than the subordinates’ assessments of the manager, the manager’s performance declined. A study of 122 Masters in Business Administration students (Korsgaard 1996) found that students who had negative self-assessments were more likely to experience declines in performance following feedback from faculty about the quality of their work on a consulting assignment than students with positive initial self-assessments. The author concludes, “When there is any sort of disagreement between self–other opinions, self-assessment may interfere with the extent to which individuals incorporate and learn from feedback.” While this could imply that self-assessment should not be used, in fact the author argues that making self-assessment more accurate should be the focus, as it would lead to more responsiveness to feedback and thus to changes in performance.

The role of feedback in the self-assessment process is particularly important if self-assessment is to be used for supervision of performance improvement. Feedback about strengths and weaknesses can, for example, provide direction for improved performance (Johnson and Cujec 1998). If feedback does lead to changes in self-assessments and behavior, it is an important tool. “The literature suggests that the extent to which feedback deviates from initial self-expectations of performance affects reactions to feedback” (Korsgaard 1996).

Steele and Ovalle (1984) believed that feedback could play a potentially vital role in mediating self-raters' abilities to more accurately simulate supervisory ratings.

Outcomes of self-assessment

The research on self-assessment has identified a number of potential benefits and limitations for both the individual and the organization. Stuart et al. (1980) state that “[R]eports in the literature have documented the efficacy of self-observation, self-recording, and self-evaluation in promoting improved performance, feelings of psychological well being, and professional development.” Additional outcomes of self-assessment on self-learning and performance improvement are discussed above.

Feedback from participants in self-assessment programs has generally been positive. Kennell et al. (1973, cited in Gordon 1992) followed medical students for two years to assess the influence of a self-assessment program. The students attributed attitudinal and behavioral changes to their self-assessment experience.

Unfortunately, little information is available on the cost of self-assessment. Gordon (1992) notes that most of the self-assessment programs he reviewed were conducted with the ordinary resources of the teaching programs, which suggests that there may not be large additional costs for self-assessment. However, Flood (1998) comments that the application of a self-assessment program in his group practice required well-developed tools and good quantitative data. The QA Project study in Indonesia of self-assessment and peer review (done to maintain provider communication skills after training) reported per-participant costs over four months: direct marginal costs of self-assessment were \$1.56 (U.S.), which increased to \$9.48 when the opportunity cost of provider and supervisor time was added. Corresponding four-month per-participant costs of peer review were \$10.98 and \$24.29 (Kim et al. 2000a, b).

Use of self-assessment in developing countries

Individual self-assessment

Documentation of individual self-assessment in developing countries is rare; our literature review yielded only two published studies.

Indonesian Midwives Association: In a 1992–93 study (MacDonald 1995), the Indonesian Midwives Association and University Research Co., LLC (URC) compared three methods for assessing midwife performance in providing family planning services in Indonesia: self-assessment, peer review, and direct observation. The assessments were imple-

mented in three cycles, with each cycle focused on a different family planning topic. From 163 to 252 self-assessment forms were completed for each cycle in the three provinces where the study was performed.

The results showed a high level of agreement among the three methods. The small differences between self-assessment and the other two methods decreased with each cycle. Nevertheless, the author concluded that direct observation with peer review is the preferred method, in part because direct observation provides additional information on how well each task was performed. Self-assessment is identified as a viable alternative for uncovering areas of weakness and is appropriate as an assessment method when direct observation is not possible. The study also compared three different formats used by the midwives to record their self-assessments and concluded that the best format was the one that asked how frequently each task was performed (always to never), rather than the ones that asked about general level of competence (very good to poor) or whether the task was performed at all (yes or no). The preferred format revealed the widest variation in responses, indicating it is most suited to identifying weaknesses in performance.

Self-assessment for performance retention: The more recent QA Project study (Kim et al. 2000a, b) measured the increase in family planning counseling performance by Indonesian midwives immediately following a one-week training and again four months later. Audiotapes of 1,210 counseling sessions were analyzed. The measurement after four months was used to evaluate the effectiveness of different retention interventions (self-assessment and peer review) in maintaining the increase in performance achieved as a result of the training. Several indicators were used to reflect performance; the most important were the increase in facilitative and informative comments by providers.

Results showed that performance improved dramatically immediately following training, but then fell off significantly at four months in the training-only group who received no retention intervention. However, the increase in facilitative comments due to training was maintained in the group receiving self-assessment plus peer review, and experienced only a small drop in the self-assessment-only group. However, the retention strategies were not successful in maintaining the increase in informative comments. The authors hypothesize that this is because the retention strategies focused on facilitative aspects and ignored informative aspects of counseling. The study concludes that the retention strategies are fairly low-cost and probably cost-effective methods for maintaining performance improvements.

COPE: Client-Oriented, Provider Efficient

COPE is a process and a set of tools that enable health clinics to incorporate self-assessment in an ongoing quality improvement program. It has been implemented in some fashion in over 35 countries (Bradley 1998). In addition to self-assessment, it provides guidance for client interviews, client-flow analysis, and developing a plan of action (Lynam et al. 1993). Staff use self-assessment checklists to assess the services they provide. The checklists address quality of medical and nursing services, staffing, community involvement, the physical facility, supplies, record keeping, organization of services, client counseling, and information and education. What makes COPE unique is its focus on the people most knowledgeable about the clinics: the staff. They evaluate their own services, identify problems, and try to develop workable solutions. The analysis involves personnel at all levels, from managerial and medical through cleaning and maintenance staff.

Unfortunately, there have been few formal assessments of COPE and no analysis of findings across project sites. Some organizations using this tool report positive experiences:

In Kenya, staff members produced an action plan identifying problems and their solutions. Of the 12 problems identified, three were addressed and two solved.

In Nepal, many problems were resolved soon after the development of an action plan based on the self-assessment.

In Ghana, Kenya, Nigeria, and Uganda, facilities reported resolving 59 percent of identified problems and 73 percent of those that could be solved without outside help (Lynam et al. 1993).

In Bangladesh, providers reported that they were more aware of and responsive to client needs and rights after implementing COPE (Stoeckel et al. 1997). They also indicated that they paid increased attention to counseling, infection prevention, and client screening after the exercise.

However, the process did not always work as envisioned. For example, in a district hospital in Kenya, implementation was derailed due to poor facilitation skills.

While more rigorous evaluations of changes in clinic performance are needed, these accounts suggest that COPE can help identify behavioral and cognitive areas requiring updating or improvement and that group self-assessment can bring about positive changes in developing country contexts.

This study also provided useful results regarding the formatting of self-assessment forms. Specifically, they began with three different formats: the first asked simply whether the subject performed certain actions, with a yes/no response; the second asked subjects to rate their competence on a four-point scale of 1 (very good) to 4 (poor); and the third how frequently certain actions were performed, on a scale of 1 (always) to 4 (never). The pilot test of these three formats showed that the first two formats did not identify weaknesses very effectively (subjects responded mostly in the affirmative in the first case and rated themselves almost always in the top in the second). The third format revealed the widest variation in responses, indicating that it was most suited to identifying weaknesses in performance.

These findings support the utility of self-assessment as a self-learning tool. Assessments were not tied to summative results or to rewards, but were a means for providers to learn how to improve their performance, and then it provided them with a framework against which to measure progress. The findings also underscore the importance of implementing self-assessment in conjunction with some other type of feedback or supervision structure, as the group that experienced self-assessment in conjunction with peer review obtained more positive outcomes. The findings are similar to those in industrialized countries, with self-assessment proving a good tool for changing specific behaviors and providing an effective format for self-learning, while benefits were more marked where some form of ongoing supervision and support was provided.

The results of additional self-assessment studies by the QA Project in Mali, Mexico, and Zimbabwe are expected in late 2001. In Mali, a study is testing whether provider self-assessment improves case management of febrile illness in children. In Mexico, an integrated program of physician self-assessment and supervisor support sought to increase retention of communication skills gained during training. Physicians used self-assessment forms to review audiotapes of their own interactions with clients. Supervisors encourage the use of the self-assessment forms, but do not ask to see the forms as this may be seen as threatening. Early results indicate a very positive reaction to the use of audiotaping for self-assessment and improved retention of supportive communication by providers in the program. In Zimbabwe, a study of the supervision system plans to use supervisor self-assessment to improve supervisory skills.

Group self-assessment

There are several notable examples of group self-assessment in international health conducted by a number of organizations, including AVSC, the International Planned

Parenthood Federation, the Population Council, and World Neighbors. AVSC's COPE (client-oriented, provider-efficient) is one of the better-documented and -formalized of these methodologies (see box). World Neighbors offers a field guide to nongovernmental organizations and community groups to undertake self-assessment (Gubbels and Koss 2000).

The QA Project has used group self-assessment in several countries, but has not measured the independent impact of group self-assessment in these activities. For example, in Nicaragua, teams trained in quality improvement undertook efforts to improve maternal and neonatal care. The teams defined several indicators of their effectiveness, such as the percentage of pregnancies for which perinatal technologies were used correctly, the percentage with a correctly used prenatal form, and client satisfaction. Each month the teams generated graphs of the indicator values and displayed the graphs on the health center bulletin board for staff, supervisors, and patients to see. The staff reported that they were continuously motivated by this self-monitoring and the progress made. While such anecdotal evidence of the positive effect of group self-assessment is available from these activities, its impact has not been systematically evaluated. The QA Project has used self-assessments by QA teams as part of a program evaluation in several countries, including Chile, Kenya, Morocco, Niger, and Russia.

Recommendations for the application of self-assessment in developing countries are largely the same as those in industrialized countries, as the available information does not allow for much interpretation in a developing country context. In light of the results obtained by Kim et al. (2000a, b), there is reason to think that self-assessment could be used in developing countries as an instrument for self-learning that can bring about changes in behavior. It is difficult to draw conclusions about its utility in changing levels of cognitive knowledge or in forming summative judgments, as there is no solid information available on these subjects. Until further work is done, it will be hard to delineate exactly which usages of self-assessment would be most practical or beneficial in a developing country context.

Research recommendations

While much research on self-assessment has been done, little has focused on its practical application, particularly in healthcare settings in developing countries. In addition, as indicated by the fact that the questions below are similar to the sections of this paper, there is little conclusive evidence regarding self-assessment and the best way to incorporate it into programs; additional research is clearly needed. Several research questions identified by the QA Project as

being of primary importance for future investigation are briefly discussed below.

Which self-assessment instruments work best?

While some authors feel that the self-assessment tool does not affect the validity of self-assessment (Harris and Schaubroeck 1988), the majority of the literature suggests that it does, and researchers continue to look for answers about which types of instrumentation are most effective (Korsgaard 1996). Knowing which self-assessment tools are most effective and for which purposes can help to guide those planning self-assessment programs. Do people prefer to respond to open-ended questions or to specific formats? Are yes/no replies or Likert scales more likely to provide useful and accurate information? Is there a maximum length that is appropriate for self-assessment tools?

Research has begun to address some of these issues. One review of self-assessment studies (Falchikov and Boud 1989) found that the number of discriminations required in a self-assessment is related to its validity, concluding that approximately 100 discriminations are associated with the best results. Mabe and West (1982) found that social comparison terminology was associated with more valid self-assessment. Falchikov and Boud (1989) concluded that the metric preferred by students, a 10-point scale, was less accurate than other scales, which suggests that there may be a trade-off to be made.

Which behaviors are more likely to be affected by self-assessment?

Many of the self-assessment studies in healthcare address efforts to improve communication skills. A few (Geissler 1973; Abrams and Kelly 1974) show its effectiveness in changing the behavior of dental students in making dental products. What this does not tell us is whether self-assessment is more effective at changing some behaviors than others. If it is to be used to effect change in the behavior of healthcare providers, knowing which behaviors it would most likely impact could guide its use.

What is the effect of self-assessment on performance?

If self-assessment is to be used as part of quality improvement/assurance efforts, it should provide information about the quality of services. If it can lead to improvements in the quality of services, it will be an even stronger tool. Another related question is: Can self-assessment be used to predict future performance?

Can self-assessment work without supervision?

The research thus far has focused on the use of self-assessment along with supervision or feedback in some form. Whether stand-alone self-assessment is plausible and sustainable is a question of importance, particularly in the developing country context. Can self-assessment work in a healthcare system where supervision is limited? Can it be implemented without initial training in self-assessment, or with limited training or on-the-job training?

What is the role of feedback?

With an increasing focus on the use of self-assessment as a supervisory tool, it is especially important to understand the effect of feedback in a self-assessment environment. Most of the feedback studies have looked at feedback that was formative rather than summative. Atwater and Yammarino (1997) note the absence of studies of the effect of feedback for evaluative or appraisal purposes. The effect of feedback in a self-assessment environment is an important consideration. If, for example, as Johnson and Ferstl (1999) suggest, the performance of under-raters declines in response to positive feedback because they are exceeding their own standards, feedback may have negative effects for some participants. Knowing what kinds of effects feedback may have and how it differs by characteristics of self-raters will be important for those implementing self-assessment programs. In addition, knowing whether or not more accurate self-assessment changes the impact of feedback would be useful as it may be that formal self-assessment increases responsiveness to feedback (Korsgaard 1996).

How should costs of self-assessment be measured?

How best to measure the costs of self-assessment has been virtually ignored. Until this question is answered, attempts to measure them will likely be inconclusive. As noted above, there are few references to the cost of self-assessment in the literature. The studies that do exist give conflicting results, in part because they include different costs of the process. One cost that has not always been included is the cost of supervision, despite the fact that almost all self-assessment includes a supervisory component. This should include time taken away from other duties to assist in feedback or coordination of self-assessment. In the case of Kim et al. (2000a, b), the cost of orienting participants to the self-assessment process and time spent on follow-up added significantly to the cost of the intervention.

Is a self-assessment program sustainable?

All of the studies found through this literature review described self-assessment programs that had been recently instituted, so they do not consider the long-term effects of self-assessment programs and skills. Issues such as whether self-assessment remains effective when conducted over long periods must be considered if it is to be used regularly. Do people continue to take self-assessments to heart when they are a standard part of their job or do they become just a form to fill out? Can students who use self-assessment transfer this skill once they are on the job?

Conclusion

A range of medical and nonmedical institutions and settings has used self-assessment for a correspondingly broad range of objectives. The means and methodologies for assessment have also been broad, most often involving a paper and pencil format and a ratings scale, but in other cases involving videotaped playback or oral self-assessment in front of peers. While different researchers have posited varying definitions of self-assessment and a few theoreticians have elaborated some pertinent theoretical underpinnings, there is a surprising dearth of theoretical literature on self-assessment—and little attempt to tie together divergent definitions and theory. Basically, self-assessment has meant a variety of things to different people with a relative lack of theoretical discussion or clarification of its underlying principles and definitions. Also, most of the research and studies to date do not delineate the theory that may inform them. This area is ripe for further work.

Self-assessment has been used as a means of summative evaluation, whose findings feed into external evaluation of a worker or student's performance, such as a grade in a class or a workplace evaluation for determining pay raises or promotions. Conversely, self-assessment has been used in a formative evaluation, designed to help the participant assess his or her performance and identify areas needing improvement, without using the findings as a final measure of one's work.

The four major uses of self-assessment are to identify learning needs, improve performance, evaluate performance, and reinforce cognitive abilities and skills after training.

Self-assessment is well suited to self-learning as a means for students or workers to learn about their own performance. However, there are some weaknesses in this realm. Much of the research is not as methodologically rigorous as the research on validity, and some important questions remain

unanswered. These include whether self-learning from self-assessment transfers from training to work contexts, what lasting impact such learning has, and the theoretical underpinnings of the self-learning process.

As with self-learning, self-assessment as an instrument for performance improvement shows promise. Some researchers feel that self-assessment functions primarily as a map of noncognitive abilities and that it may be effective in altering them. Certainly, there are more examples from the medical literature showing its use for noncognitive than cognitive abilities, such as in changing medical practitioners' communications skills. However, several important aspects of self-assessment's impact on behavior remain underresearched, including the durability of newly acquired behaviors, behaviors most effectively influenced by self-assessment, and the theoretical mechanisms whereby behavior changes. Despite these research gaps, self-assessment deserves further use as a tool for performance improvement.

The primary focus of research on self-assessment's role in performance appraisal has been on validity: how well self-evaluations correlate with assessments from more objective sources. On average, validity appears moderate to low in both the medical and nonmedical literature (although the bulk of the medical literature comes from medical schooling, not clinical, contexts). In general, the nonmedical academic and workplace literature shows that subjects overestimate their own performance, whereas the medical literature, comprising primarily work on health profession students, shows that medical students tend to underestimate their own performance. It is plausible that this tendency to underestimate performance may also extend to the medical workplace, although this needs investigation.

Some research has examined factors that may influence validity, ranging from personality variables, such as self-esteem and narcissism, to differences in the format of the measurement instrument and the setting for the evaluation (for example, conditions of anonymity versus disclosure of self-assessment). While the research has identified some factors that may improve validity, such as guarantees of anonymity and practice at self-assessing, this area needs further work, and the potential to strengthen validity remains questionable. For these reasons, self-assessment in a summative monitoring and evaluation context, or in a selection context, is not optimal.

The published literature suggests that effective self-assessment requires supervision and ongoing support. The literature does not yield any examples of self-assessment conducted without supervision. Supervision can be costly,

and in fact may be the largest component in the cost of self-assessment. It may be difficult to involve providers working without supervision in quality assurance through self-assessment; they may need some structure to support them.

There are several other factors that may improve self-assessment. Although the evidence is mixed, there is support for the idea that self-assessment improves with practice, like a skill that must be learned and honed over time. The evidence also suggests that self-assessment works best where the behaviors targeted are specific rather than general and where criteria for evaluation are clearly delineated, defined, and discussed beforehand. It seems helpful to guarantee anonymity to improve validity. Regarding format, scales of 100 metrics appear optimal, and instruments should not be overly long. On the whole, though, the question of format requires further research. While some of the factors outlined here provide a modicum of direction, more research needs to focus on strengthening both the impact and validity of self-assessments.

The potential benefits from using self-assessment may play an important role in developing country health sector quality assurance. It may be a lower-cost evaluation mechanism than many others and relatively easy to implement. Because it is linked to self-direction, it may also be the most appropriate tool for adult learners. There are reports that it can enhance self-esteem, give participants greater ownership over the assessment process, and improve communications between supervisors and other staff. While significant work still needs to be done to determine the contexts and methods through which self-assessment will have the most impact, it is potentially a valid tool for self-evaluation that deserves further utilization and attention.

Table 4 ■ Studies of Validity

Author/Date	Target Group	Target Area/Topic	Study Methodology
Arnold et al. (1985)	US Baccalaureate-medical students (n=211): clinical skills in docent rotations in internal medicine	Comparison of self-assessment to ratings by docents, basic scientists, and faculty, as well as to grades and quarterly and national exams	This longitudinal study was designed to look at the development of self-assessment skills over time.
Das et al. (1998)	First-year medical students (n=64) in the United Arab Emirates	Self- and tutor evaluations in problem-based learning (PBL) tutorials: also compared to scores on a modified essay question exam	Students self-assessed their performance in PBL tutorials over a two-year period. Tutors also assessed each student.
Church (1997)	Senior service providers in a business advisory and professional services firm (n=152)	Assessments of management behavior and performance outcomes by self, direct reports, peers, supervisors, and clients	Summary scores for the raters were used to look at the congruence between self- and other assessments.
Farh and Dobbins (1989a)	Undergraduate students (n=163)	Looked at whether or not social comparison information leads to greater discrepancy between self- and supervisor ratings	This experimental study assigned students to work groups. Editorial tasks were completed by 5 group members with knowledge that the supervisor would see their self-evaluations of their effectiveness. The social comparison group was allowed to review the work of their coworkers while the control group was not. A head editor in each group evaluated the work of subordinates. Indicators of actual performance were used as a gold standard.
Frye et al. (1991)	Medical students (n=22)	Compared student self-assessments of performance on exam questions with actual performance	This longitudinal study compared the accuracy of self-assessments across four examinations.
Henbest and Fehrsen (1985)	Medical students (n=19) in South Africa	Compared self-assessment at the end of a clinical rotation in family medicine to those of a tutor and head of the department	Students and faculty scored student performance using an evaluation developed by students.
Johnson and Cujec (1998)	Canadian, intensive care unit residents (n=60)	Compared physician, nurse, and self-ratings on a Global Rating Scale (GRS) and a Behaviorally Anchored Rating Scale (BARS): multiple-choice test also used for comparison	In this prospective cohort study, medical residents were assessed by selves, 3 physicians, and 6 nurses.
Love and Hughs (1994)	Police officers (n=73)	Compared self-assessments of job performance with scores on a promotional examination	Candidates for promotion submitted their self-assessment a week before the written exam.
Nowack (1992)	Entry- to mid-level managers in large organizations (n=335)	Compared self-assessments to those of up to 5 other people (peers, subordinates, and superiors)	Managers and 5 other people suggested by the manager completed a standard Management Practices Questionnaire (MPQ).

Statistically Significant Relationships	Other Findings	Author/Date
Most significant correlate of self-ratings was other self-ratings by the student. Most significant predictor of self-ratings was docent's rating in the prior year (Pearson's correlation: 0.25). Grades and scores on medical school exams and national exam were not correlated with self-assessment.	Students who had higher GPAs, higher quarterly and national examination scores, and superior faculty ratings were more likely to rate themselves lower than did the docents.	Arnold et al. (1985)
In 1994, there was a statistically significant difference in the mean scores of tutors and students on responsibility. In 1995, there was a difference on self-awareness of the student. Overall responsibility scores and critical analysis scores were also correlated. No correlation was found between ratings by either students or tutors and end-of-year exam scores.	High achievers were less likely to give themselves a high rating than low achievers.	Das et al. (1998)
Correlations were low for behavioral and performance indicators. Self- and supervisor ratings were the most highly correlated. Ratings by different types of raters within the organization were highly significant ($r = .27-.36$); correlations with client ratings were lower. Correlations differed by type of observer.	Providers rated themselves higher than direct reports, peers, or supervisors. Clients also gave higher ratings.	Church (1997)
Participants in the social comparison group were significantly more accurate in evaluating their effectiveness compared to the performance indicators than the control group (correlation was .51 in the social comparison group, and .29 in the control). Correlation between self- and supervisor ratings was higher in the social comparison group (.42) than in the control group (.13). These results were only significant for the overall rating of effectiveness and not for the individual indicators.		Farh and Dobbins (1989a)
Correlations between student self-assessments and expert scores were not significant (with two exceptions). Self-assessment accuracy improved between the first 3 exams but declined on the last.	Students were more likely to overestimate than underestimate their performance; however, overestimations steadily declined across the 4 exams.	Frye et al. (1991)
Scores of students were significantly correlated ($0.74, p < .01$) with those of faculty members.		Henbest and Fehrsen (1985)
The means, medians, and quartiles for the GRS and BARS were roughly equivalent. The scores of physicians were correlated with the multiple-choice test scores. The correlations between physician and nurse scores were significant except on humanistic qualities. There was a significant correlation between nurse and self-scores for procedural skills. Medical knowledge scores of physicians were correlated with the self-evaluations.		Johnson and Cujec (1998)
Higher self-assessment scores correlated with higher exam scores on 5 self-assessment factors.	A majority of the candidates were not in favor of the use of self-assessment in the promotion process.	Love and Hughs (1994)
Across 19/20 MPQ scales, the mean scores of others were significantly different than the self-score, with managers reporting all behaviors more frequently than did those reporting on them. Managers had inflated perceptions of how frequently they practiced certain behaviors than did others. Correlations between self- and other ratings were low, ranging from 0.12–0.30.		Nowack (1992)

Table 4 ■ Studies of Validity (Continued)

Author/Date	Target Group	Target Area/Topic	Study Methodology
Regehr et al. (1996)	Third-year clinical clerks on a psychiatry rotation (n=25)	Compared self-assessment of interview skills with ratings by experts	Students were oriented to the self-assessment checklist before interviewing a standardized patient. Two experts watched the interaction and rated the interview. After the interview, the student completed the self-assessment checklist.
Stuart et al. (1980)	US medical residents (n=56)	Compared student self-assessments of patient interviews with faculty assessments	Two patient interviews by each resident were videotaped. Both the resident and 3 faculty members reviewed the tape separately and rated the residents' skills. Mean scores from the faculty members were compared to the resident scores.
Sullivan et al. (1999)	Third-year medical students in surgical clerkship (n=154)	Compared self-, peer, and tutor ratings of performance in problem-based tutorials	Faculty and students provided 3 global ratings of student performance after 6 meetings to discuss 3 cases.
Thornton (1968)	Managers in a large US manufacturing corporation (n=64)	Compared performance appraisal ratings from a supervisor to the self-appraisal	The manager and his immediate supervisor or superior completed a 27-item appraisal form aimed at establishing the basis for development.
Woolliscroft et al. (1993)	Third-year US medical students (n=142)	Compared clinical self-assessments to external performance measures (grade point averages, standardized exam scores, and ratings by faculty and residents)	Self-assessments were completed at the beginning and the end of clinical rotations. Faculty and resident ratings were mean ratings assigned by all faculty/residents over the course of the 12-week rotation.

Statistically Significant Relationships	Other Findings	Author/Date
The mean correlation between the student's ranking of her skills with that of the experts was .43 ($p < .001$). The correlation between the student self-assessment score and the experts' overall rating of performance was not significant (0.19, $p > .10$).		Regehr et al. (1996)
There were moderate but significant correlations on three skills for which criteria was more specific (closure, response to patient, and therapy and disposition).	Only 35% of the variability in reviewer scores represented agreement between the raters.	Stuart et al. (1980)
Highest correlations between peer and faculty ratings: varied by performance group ($r = .50$ for independent learning, $.54$ for group participation, and $.24$ for problem solving; $p < .01$). Lowest correlations were between self- and faculty ratings ($.24$, $.18$, and $.11$, respectively).	Overall, proportion of variance explained was small (most less than 6%).	Sullivan et al. (1999)
Overall, mean scores of the managers were higher than those of the supervisors. The supervisors and managers did not agree on the areas in which performance was satisfactory and unsatisfactory.		Thornton (1968)
No correlation was found between prior external assessments (e.g., MCAT) and self-assessments. Students in the lowest quartile on external assessments ranked themselves higher than other students. Some weak but significant correlations existed between student assessments at the end of the rotation and faculty ratings.		Woolliscroft et al. (1993)

References

- AAP (American Academy of Pediatrics). 2000. Ambulatory Care Quality Improvement Program (ACQIP). <<http://www.aap.org/prof/acqip.htm>>.
- Abrams, R.G., and M.L. Kelley. 1974. Student self-evaluation in a pediatric-operative technique course. *Journal of Dental Education* 38(7):385–91.
- Adamow, C. 1982. Self-assessment: A quality assurance tool. *Journal of the American Dietetic Association* 81(1):62–63.
- Adams, W.R., T.H. Ham, B.H. Mawardi. 1974. Research in self-evaluation for clinical teachers. *Journal of Medical Education* 49:1166.
- AIHA (American International Health Alliance). 2000. *A report on the self-assessment of partnerships in the newly independent states and central and eastern Europe*. <<http://www.aiha.com/english/partners/assess/>>.
- Arnold, L., T.L. Willoughby, and E.V. Calkins. 1985. Self-evaluation in undergraduate medical education: A longitudinal perspective. *Journal of Medical Education* 60(1):21–28.
- Arthur, H. 1995. Student self-evaluations: How useful? How valid? *International Journal of Nursing Studies* 32(3):271–76.
- ASHP (American Society of Health-Systems Pharmacists). 2000. Pharmacy directors to receive ASHP-endorsed tool for assessing medication-use safety in hospitals. <http://www.ashp.org/public/news/breaking/pharm_dir.html>.
- Atwater, L.E., and E.J. Yammarino. 1997. Self-other rating agreement: A review and model. In *Research in Personnel and Human Resource Management*, G.R. Ferris, ed. Greenwich, CT: Jai Press, Inc.
- Ballenger, C.R. 1974. Trust as a basis for course evaluation: An alternative to teacher judgements. *Science Teacher* April:33–34.
- Bandura, A. 1977. Self-efficacy: Towards a unifying theory of behavioural change. *Psychological Review* 84(2):191–215.
- Best, M., R.J.B. Carswell, and S.D. Abott. 1990. Self-evaluation for nursing students. *Nursing Outlook* 38(4):172–77.
- Bradley, J. 1998. Using COPE to improve quality of care: The experience of the Family Planning Association of Kenya. *Quality/Calidad/Qualite* No. 9, New York, NY: The Population Council.
- Brew, A. 1999. Towards autonomous assessment: Using self-assessment and peer-assessment. In *Assessment Matters in Higher Education: Choosing and Using Diverse Approaches*, S. Brown and A. Glasner, eds. Buckingham, U.K.: Society for Research into Higher Education.
- Cheung, G.W. 1999. Multifaceted conceptions of self-other ratings disagreement. *Personnel Psychology* 52(1):1–36.
- Church, A.H. 1997. Do you see what I see? An exploration of congruence in ratings from multiple perspectives. *Journal of Applied Social Psychology* 27(11):983–1020.
- Cochran, S.B., and M.C. Spears. 1980. Self-assessment and instructors' ratings: A comparison. *Journal of the American Dietetic Association* 76:253–55.
- Das, M., D. Mpofo, E. Dunn, and J.H. Lanphear. 1998. Self and tutor evaluations in problem-based learning tutorials: Is there a relationship? *Medical Education* 32(4):411–18.
- Falchikov, N., and D. Boud. 1989. Student self-assessment in higher education: A meta-analysis. *Review of Educational Research* 59(4):395–430.
- Farh, J.L., and G.H. Dobbins. 1989a. Effects of comparative performance information on the accuracy of self-ratings and agreement between self- and supervisor ratings. *Journal of Applied Psychology* 74(4):606–10.
- Farh, J.L., and G.H. Dobbins. 1989b. Effects of self-esteem on leniency bias in self-reports of performance: A structural equation model analysis. *Personnel Psychology* 42:835–49.
- Festinger, L. 1954. A theory of social comparison processes. *Human Relations* 7:117–40.
- Fincher, R.E., L.A. Lewis, and T.T. Kuske. 1993. Relationships of interns' performances to their self-assessments of their preparedness for internship and to their academic performances in medical school. *Academic Medicine* 68(2):S47–S50.
- Flood, S.C. 1998. Using qualitative self-evaluation in rating physician performance. *Family Practice Management* 5(5): 22–24, 27–30, 33–34.
- Frye, A.W., B.F. Richards, E.W. Bradley, and J.R. Philp. 1991. The consistency of students' self-assessments in short-essay subject matter examinations. *Medical Education* 25:310–16.
- Fuhrmann, B.S., and M.J. Weissburg. 1978. Self-assessments. In *Evaluating Clinical Competence in Health Professionals*, M.K. Morgan and D.M. Irby, eds. St. Louis: C.V. Mosby Co.
- Geissler, P.R. 1973. Student self-assessment in dental technology. *Journal of Dental Education* 37(9):19–21.
- Gordon, M.J. 1991. A review of the validity and accuracy of self-assessments in health professions training. *Academic Medicine* 66(12):762–69.
- Gordon, M.J. 1992. Self-assessment programs and their implications for health professions training. *Academic Medicine* 67(10):672–79.
- Gordon, M.J. 1997. Cutting the Gordian knot: A two-part approach to the evaluation and professional development of residents. *Academic Medicine* 72(10):876–80.
- Greenwald, A.G. 1980. The totalitarian ego: Fabrication and revision of personal history. *American Psychologist* 35:603–18.
- Gubbels, P., and C. Koss. 2000. *From the Roots Up: Strengthening Organizational Capacity through Guided Self-Assessment* (2nd edition). Oklahoma City: World Neighbors.
- Harris, M.M., and J. Schaubroeck. 1988. A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. *Personnel Psychology* 41:43–62.

- Hay, J.A. 1995. Investigating the development of self-evaluation skills in a problem-based tutorial course. *Academic Medicine* 70(8):733–35.
- Henbest, R.J., and G.S. Fehrsen. 1985. Preliminary study at the medical university of Southern Africa on student self-assessment as a means of evaluation. *Journal of Medical Education* 60(1):66–68.
- Heneman, H.G. 1980. Self-assessment: A critical analysis. *Personnel Psychology* 33:297–300.
- Hinsz, V.B., and D.C. Matz. 1997. Self-evaluations involved in goal setting and task performance. *Social Behavior & Personality* 25(2):177–82.
- Hughes, R.E., K.G. Bass, and E.J. Hebert. 1997. Self-evaluation with and without feedback as a predictor of attributions. *Psychological Reports* 80:1075–85.
- Hunter, J.E., and R.F. Hunter. 1984. Validity and utility of alternative predictors of job performance. *Psychological Bulletin* 96(1):72–98.
- Jackson, S. 1998. Europe supports UK government in putting quality at the heart of healthcare. *Health Manpower Management* 24(3):100–03.
- Jackson, S. 1999. Achieving a culture of continuous improvement by adopting the principles of self-assessment and business excellence. *International Journal of Healthcare Quality Assurance* 12(2):59–64.
- John, O.P., and R.W. Robins. 1994. Accuracy and bias in self-perception: Individual differences in self-enhancement and the role of narcissism. *Journal of Personality and Social Psychology* 66(1):206–19.
- Johnson, D., and B. Cujec. 1998. Comparison of self, nurse, and physician assessments of residents rotating through an intensive care unit. *Critical Care Medicine* 26(11):1811–16.
- Johnson, J.W., and K.L. Ferstl. 1999. The effects of interrater and self–other agreement on performance improvement following upward feedback. *Personnel Psychology* 52(2):271–303.
- Kaiser, S., and J.J. Bauer. 1995. Checklist self-evaluation in a standardized patient exercise. *The American Journal of Surgery* 169(4):418–20.
- Kak, N., B. Burkhalter, and M. Cooper. 2001. Measuring the competence of healthcare providers. *Operations Research Issue Paper* 2(1). Bethesda, MD: Published for the U.S. Agency for International Development (USAID) by the Quality Assurance Project.
- Katz, R. 1970. *Self-assessment workshop (SAW): Final research report*. Peace Corps Contract #80-1531, Task Order 3. Cambridge, MA: Human Development Foundation.
- Kennell, J.M., C.R. Tempio, and M.Z. Wile. 1973. Self-evaluation by first-year medical students in a clinical science program. *British Journal of Medical Education* 7:230–38.
- Kernis, M.H., J. Jadrich, P. Gibert, and C.R. Sun. 1996. Stable and unstable components of self-evaluations: Individual differences in self-appraisal responsiveness to feedback. *Journal of Social and Clinical Psychology* 15(4):403–48.
- Kim, Y.M., F. Putjuk, E. Basuki, and A. Kols. 2000a. Self-assessment and peer review: Improving Indonesian service providers' communication with clients. *International Family Planning Perspectives* 26(1):4–12.
- Kim, Y.M., F. Putjuk, A. Kols, and E. Basuki. 2000b. Improving provider-client communication: Reinforcing IPC/C training in Indonesia with self-assessment and peer review. *Operations Research Results* 1(6). Bethesda, MD: Published for the U.S. Agency for International Development (USAID) by the Quality Assurance Project.
- Klein, W.M. 1997. Objective standards are not enough: Affective, self-evaluative, and behavioral responses to social comparison information. *Journal of Personality and Social Psychology* 72(4):763–74.
- Klimoski, R., and L. Inks. 1990. Accountability forces in performance appraisal. *Organizational Behavior and Human Decision Processes* 45:194–208.
- Korman, A.K. 1970. Toward a hypothesis of work behavior. *Journal of Applied Psychology* 54:31–41.
- Korsgaard, M.A. 1996. The impact of self-appraisals on reactions to feedback from others: The role of self-enhancement and self-consistency concerns. *Journal of Organizational Behavior* 17(4):301–11.
- Kruger, J., and D. Dunning. 1999. Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology* 77(6):1121–34.
- Levine, E.L. 1980. Introductory remarks for the symposium "Organization applications of self-appraisal and self-assessment: Another look." *Personnel Psychology* 33:259.
- Lichtenstein, S., and B. Fischhoff. 1977. Do those who know more also know more about how much they know? *Organizational Behavior and Human Performance* 20:159–83.
- Love, K.G., and F.V. Hughes. 1994. Relationship of self-assessment ratings and written test score: Implications for law enforcement promotional systems. *Public Personnel Management* 23(1):19–29.
- Lux, M. 2000. Personal communication.
- Lynam, P., L.M. Rabinovitz, and M. Shobowale. 1993. Using self-assessment to improve the quality of family planning clinic services. *Studies in Family Planning* 24(4):252–60.
- Mabe, P.A., and S.G. West. 1982. Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology* 67(3):280–96.
- MacDonald, P. 1995. The peer review program of the Indonesian Midwives Association (Final Report of Phase Two of the Pilot Project). Bethesda, MD: University Research Co., LLC and the Indonesian Midwives Association.
- Maguire, P. 1990. Can communication skills be taught? *British Journal of Hospital Medicine* 43(3):215–16.

- Makiney J.D., and P.E. Levy. 1998. The influence of self-ratings versus peer ratings on supervisors' performance judgments. *Organizational Behavior and Human Decision Processes* 74(3):212–28.
- Marienau, C. 1999. Self-assessment at work: Outcomes of adult learners' reflections on practice. *Adult Education Quarterly* 49(3):135–46.
- Marquez, L. In press. Helping healthcare providers perform according to standards. *Operations Report Issue Paper*. Bethesda, MD: To be published for the U.S. Agency for International Development (USAID) by the Quality Assurance Project.
- Martin, D., G. Regehr, B. Hodges, and N. McNaughton. 1998. Using videotaped benchmarks to improve the self-assessment ability of family practice residents. *Academic Medicine* 73(11):1201–06.
- Mason, J.L., S.E. Barkley, M.M. Kappelman, D.E. Carter, and W.V. Beachy. 1988. Evaluation of a self-instructional method for improving doctor-patient communication. *Journal of Medical Education* 63(8):629–35.
- Mayall, D., and M. Maze. 1985. Toward a taxonomy of self-assessable skills. *Journal of Career Development* 11(4):263–71.
- Morton, J.B., and W.A.A. Macbeth. 1977. Correlations between staff, peer and self-assessments of the fourth-year students in surgery. *Medical Education* 11(3):167–70.
- Myhre, H.O. 1998. Critical self-assessment—Is it too much for a surgeon? *Annales Chirurgiae et Gynaecologiae* 87(2):124–25.
- Nowack, K.M. 1992. Self-assessment and rater-assessment as a dimension of management development. *Human Resources Development Quarterly* 3(2):141–55.
- Palmer, P.B., J.N. Henry, and D.A. Rohe. 1985. Effect of videotape replay on the quality and accuracy of student self-assessment. *Physical Therapy* 65(4):497–501.
- Parboosingh, J. 1998. Role of self-assessment in identification of learning needs. *Journal of Continuing Education in the Health Professions* 18(4):213–19.
- Pitt, D.J. 1999. Improving performance through self-assessment. *International Journal of Healthcare Quality Assurance* 12(2):45–53.
- Pym, D.L.A., and H.D. Auld. 1965. The self-rating as a measure of employee satisfactoriness. *Journal of Occupational Psychology* 39:103–13.
- Regehr, G., B. Hodges, R. Tiberius, and J. Lofchy. 1996. Measuring self-assessment skills: An innovative relative ranking model. *Academic Medicine* 71(10):S52–S54.
- Reid, B. 1993. But we're already doing it! Exploring a response to the concept of reflective practice in order to improve its facilitation. *Nurse Education Today* 14(2):82–90.
- Reilly, R.R., and G.T. Chao. 1982. Validity and fairness of some alternative employee-selection procedures. *Personnel Psychology* 35:1–62.
- Risucci, D.A., A.J. Tortolai, and R.J. Ward. 1989. Ratings of surgical residents by self, supervisors and peers. *Surgical Gynecology and Obstetrics* 169(6):519–26.
- Rooke, T.R. 1998. A critical review of reflective practice in its role towards personal and professional development. *Journal of the Royal Naval Medical Service* 84(2):66–70.
- Rosato, F.E. 1972. Self-assessment and continuing medical education. *Surgery* 71(4):642–44.
- Sclabassi, S.E., and S.K. Woelfel. 1984. Development of self-assessment skills in medical students. *Medical Education* 18(4):226–31.
- Sedikides, C. 1993. Assessment, enhancement, and verification determinants of the self-evaluation process. *Journal of Personality and Social Psychology* 65(2):317–38.
- Shore, L.M., L.E. Tetrick, and T.H. Shore. 1998. A comparison of self-, peer, and assessor evaluations of managerial potential. *Journal of Social Behavior and Personality* 13(1):85–101.
- Shrauger, J.S., and T. Osberg. 1981. The relative accuracy of self-predictions and judgments by others in psychological assessment. *Psychological Bulletin* 90:322–51.
- Sideris, D.A., P. Tsouna-Hadjis, S. Toumanidis, and S.D. Mouloupoulos. 1990. A self-learning approach to history taking. *Medical Education* 24(1):46–51.
- Stackhouse, J., and A. Furnham. 1983. A student-centered approach to the evaluation of clinical skills. *British Journal of Disorders Communication* 18(3):171–79.
- Stahl, T. 1998. Self-Assessment: A royal road to quality assurance for continuing training? *Vocational Training: European Journal* 15:33–45.
- Steele, R.P., and N.K. Ovalle. 1984. Self-appraisal based upon supervisory feedback. *Personnel Psychology* 37:667–85.
- Stoekel, Jr., Z.A. Hakim, S. Sarker, and H. Begum. 1997. *Evaluation report of the Systems Approach Project*. Bangladesh: AVSC International.
- Stuart, M.R., H.S. Goldstein, and F.C. Snope. 1980. Self-evaluation by residents in family medicine. *The Journal of Family Practice* 10(4):639–42.
- Sullivan, M.E., M.A. Hitchcock, and G.L. Dunnington. 1999. Peer and self-assessment during problem-based tutorials. *American Journal of Surgery* 177(3):266–69.
- Taylor, J.A. 1998. Self test: A flexible self-assessment package for distance and other learners. *Computers & Education* 31(3):319–28.
- Thornton, G.C. 1968. The relationship between supervisor and self-appraisals of executive performance. *Personnel Psychology* 21:441–55.
- Thornton, G.C. 1980. Psychometric properties of self-appraisals of job performance. *Personnel Psychology* 33:263–71.

Trevitt, C., and M. Pettigrove. 1995. Towards autonomous criterion-referenced assessment and self-assessment: A case study. Invited keynote paper at 2nd European Electronic Conference on Assessment and Evaluation: Recent and Future Developments, Feb. 6–10. European Association for Research into Learning and Instruction, SIG Assessment and Evaluation List. Administered via <listserv@nic.surfnet.nl>.

Trope, Y., and E.M. Pomerantz. 1998. Resolving conflicts among self-evaluative motives: Positive experiences as a resource for overcoming defensiveness. *Motivation and Emotion* 22(1): 53–72.

Wells, L.E., and P.D. Sweeny. 1986. A test of three models of bias in self-assessment. *Social Psychology Quarterly* 49(1):1–10.

Williams, W.E., and D.A. Seiler. 1973. Relationship between measures of effort and performance. *Journal of Applied Psychology* 57:49–54.

Woolliscroft, J.O., J. TenHaken, J. Smith, and J.G. Calhoun. 1993. Medical students' clinical self-assessments: Comparisons with external measures of performance and the students' overall self-assessments of overall performance and effort. *Academic Medicine* 68(4):285–94.

Zabarenko, R., and L. Zabarenko. 1978. *The Doctor Tree*. Pittsburgh: University of Pittsburgh.

Continued from inside front cover

Operations Research Issue Paper
Operations Research Issue Papers present important background information about key subjects relevant to the QA Project's technical assistance. Each paper reviews current research (both published and unpublished, theoretical and operational) on a subject and makes recommendations for further research and productive lines of inquiry for the project's technical staff, external researchers, and health professionals.

Acknowledgement

This paper was edited by Bart Burkhalter, Sue Brechin, and Beth Goodrich.

Quality Assurance Project
7200 Wisconsin Avenue, Suite 600
Bethesda, MD 20814-4811
Tel: 301-941-8550
Fax: 301-941-8427
E-mail: qapdissem@urc-chs.com
www.qaproject.org