Journal of Health & Population in Developing Countries

Defining the Components of "Performance" in Family Planning/Reproductive Health for Conceptual Clarity and Sounder Measurement

Alfredo L. Fort
Constance J. Newman

Abstract

Within the family planning/reproductive health literature, the word "performance" is used to refer to different concepts. We attempt to standardize its definition and operationalize its measurement among international assistance organizations. In the field of Human Performance Technology, the definition of performance is both behavior –or activities and the accomplishments produced by such behaviors / activities. We describe typical methodologies used to measure each component as well as particular challenge facing the measurement and reporting of "performance".

Key Words – Provider performance; family planning; reproductive health; performance measurement; behavior; accomplishments; client satisfaction; quality of care; client provider interaction.

Address for correspondence: Alfredo L. Fort, Director of Monitoring and Evaluation,
The PRIME II project at Intrah/UNC School of Medicine, 1700 Airport Road, Chapel Hill, NC 27599-8100.
Email: afort@intrah.org Telephone: (919) 962-0922 Fax: (919) 962-7178

The ***Journal of Health and Population in Developing Countries (ISSN 1095-8940)***
is a publication of the Department of Health Policy and Administration, School of Public Health
University of North Carolina at Chapel Hill.

UNC SCHOOL OF PUBLIC HEALTH
DEPARTMENT OF HEALTH POLICY AND ADMINISTRATION

Introduction

The family planning/reproductive health (FP/RH) field has used concepts and methodologies from business and industry (e.g. Focus Group Research, Quality Assurance) to improve the implementation and evaluation of its programs. Recent years have also seen an increasing use of terminology and conceptual approaches from the field of Human Performance Technology (HPT) for human resource development (HRD)-related initiatives in FP/RH programs. Performance-related approaches such as Performance Improvement are gaining ground among international development assistance organizations (McCaffery et al., 1999; Luoma et al., 2000; JHPIEGO, 2001; de la Peza, 2001; Ashton, 2001; Mielke, 2001). Donor agencies such as USAID are increasingly promoting fora and tools based on these approaches (e.g. Performance Improvement Consultative Group, Performance Monitoring Plans). The area of health worker evaluation has been especially influenced by performance approaches, encompassing initiatives that assessed competence and learning at training sites, the transfer of learning to the job and performance assessment at the workplace **(**Kirkpatrick, 1998; Robinson & Robinson, 1990; PRIME, 1997; Newman, 2002; Dohlie, 2001; Broad, 1997). In fact, the effects of training and other supportive actions on provider performance have become commonplace in RH/FP research and evaluation.

However, this widespread interest in performance has not been accompanied by a standard definition of performance. Consequences of this lack of standardization include ambiguity and confusion with obvious implications for measurement. In this paper, we present definitions of performance from the HPT literature, and review the use of the term *performance* by HRD or service-related FP/RH agencies working in international assistance. We illustrate how the term *performance* is used to denote a variety of constructs in the FP/RH field. We also discuss the issues associated with developing and using a standard definition in the FP/RH field, a field that presents unique medical, social and political challenges.

What is Performance?

According to experts in the field of Human Performance Technology (HPT), performance is a summary term used to indicate behavior or activities and the accomplishments produced by such behavior/activities (Gilbert, 1978; Harless, 1992; Dean, 1994). Behavior refers to actions a person takes to produce a result, and accomplishments refer to the results of such behavior that have value to an organizational goal. Judging performance may also involve assessing the time frame in which a job task was carried out. These two components, behavior and accomplishments, should be causally linked. That is, there has to be a reasonable assumption that specific behaviors are sequentially related to measurable outputs or outcomes. The measurement of behaviors has been linked to the measurement of job-related accomplishments in the performance literature (Robinson & Robinson, 1990; Lindsley, 1999; Kaufman, 1998). In a number of cases, these linkages are obvious, based on well-established research criteria and experience. However, often demonstrating this link becomes a challenging exercise, since the two components stand at different levels of measurement: Behavior will most often be assessed at the individual level, while the results of behavior tend to capture wider effects, at system or organizational levels. Moreover, in many instances a work-related accomplishment may be a result of simultaneous interventions, involving multiple concomitant factors, thus making the analysis of causal relationships and attribution extremely difficult. This aspect will be reviewed later in more detail.

Measuring Performance in FP/RH

Measuring Behavior

Measuring "performance" often refers to measuring service provider behavior in the RH/FP field. Behavior can be measured through a variety of techniques. The simplest is based on self-assessment tools, which the performers use to check and/or describe whether or not they carry out specific job tasks,

sometimes rating their own proficiency. These methods can be useful for empowering workers to improve their capabilities, especially if conducted in a structured way over time. However, these methods are regarded as subjective in nature and less reliable from a methodological standpoint.

Behavior can also be measured through observations of on-the-job tasks by trained observers or experts, which add a higher level of objectivity, but are not free of methodological problems such as observers' influence over workers' behavior in carrying out specific tasks or interacting with clients. A better design (though costly to implement) is to train and deploy "mystery" or simulated clients as observers and raters (Stillman & Gillers, 1986; Huntington et al., 1990; Population Council, 1998). Behavior can be recorded as to whether or not expected tasks were executed (through "Yes/No" responses) or by rating the quality of each task in relation to a standard, using a continuum of ordered categories (e.g. a 5-point scale from "Excellent" to "Unacceptable"). In either case, assessors can analyze indicators separately to get analytic scores, or add up indicators to make holistic scores. Analytic scores are useful in identifying specific areas of strength and weakness (e.g. minimum clinical skills required for IUD insertion), while holistic scores can provide a summary measure of the "quality" or average behavior of a group of providers. Standards can be set for providers to pass minimum "thresholds." Organizations can decide on "critical tasks" (such as those associated with infection prevention) or elements without the passing of which a service provider will not be considered to have met a behavioral standard (Carlyle & Ellison, 1984).

Measuring behavior can also include costs of carrying out tasks and the overall job (Dean & Ripley, 1997; Brethower, 2001). Costing adds an "efficiency" dimension to the measurement by showing the accomplishment/cost ratio. Costing and cost-effectiveness analyses may help determine whether specific tasks are required at all or need to be restructured or modified.

Measuring Accomplishments

Measuring "performance" also refers to measuring the service provider's work-related *accomplishments* in the RH/FP field. Quantifying and qualifying the consequences and outputs of behavior thereby constitute the second major component in performance measurement. Indicators related to work-related accomplishments should be agreed to in advance, and a determination made about their relation or attribution to specific behaviors. For example, before measurement begins it must be clear in what way a worker is responsible for a specific product or outcome. For example, in the case of an assembly of parts, to what extent does an individual worker's behavior contribute to the overall production of a unit? In the field of FP/RH program evaluation, accomplishments derived from provider behavior may relate to the number and characteristics of clients (e.g. adolescents, men/women, single/in union clients) who attended a clinic by service, or the number of home visits made to deliver preventive services. Client records and clinic statistics, complemented with provider interviews, are the most common methods to elicit such information. Though posing some interpretation difficulties, such outcomes can be complemented with measures of client satisfaction with services. Other measures of accomplishments include the prevalence of use of services, obtained through household surveys. However, as will be discussed below, there are methodological issues arising in attributing clinic- or community-based outcomes to the behavior of individual workers.

Challenges to the measurement and reporting of performance in FP/RH programs

After reviewing the two-component definition of performance from the field of HPT and the likely sources of data for these two components, it might appear to be a simple task to measure and report on service provider performance. However, measuring and reporting on these two components is far from universal practice. In addition, measuring performance in the context of FP/RH programs has unique challenges. These issues will be reviewed in the next section.

Behavior as "Performance"

A common practice, inconsistent with a full definition of performance, has been to equate the measurement of behavior alone as *job performance*. This practice has been documented in reports from organizations dealing with the training and support of providers. Sometimes, it is unclear what aspects of behavior are being measured, such as the *proficiency* or the *competence* of trained providers to complete tasks, or the provider's knowledge and skills. In part, the use of the term "performance" to denote behavior comes from its commonsense connotation: How the service provider is *performing*. For example, the expression "performing to standard" has often meant simply doing something "the right [accepted] way," and has been used as a synonym for *quality* (Nicholas, 2001), i.e. a behavioral connotation, without regard for the outcomes of such behavior.

The increased use of facility surveys has extended the employment of *performance* in different ways. Using the quality of care (QoC) paradigm, clinic inventory measures (e.g. availability of supplies), service structure (e.g. privacy) and worksite dynamics (e.g. clients' waiting times) have been combined with providers' behavior, and have been integrated in a variety of tools, such as the *QIQ* and *SPA* instruments (MEASURE *Evaluation,* 2001; Bessinger, 2001; Fronczak, 2001). Thus, it is not unusual to read about individuals' or service delivery points' (SDPs) *performance* when referring to a combination of provider behavior and other measures of quality of care.

Several things explain the widespread use of behavior to report on performance. First, a worker's behavior is linked more closely with project design and intervention phases, and this facilitates the interpretation of evaluation findings. Second, the accessibility of the units of measurement and the relative simplicity of their measurement are advantages not to be underestimated. For example, observation and measurement of provider competence or client-provider interaction (CPI), though not without problems (e.g. the "Hawthorne effect" mentioned earlier, the need to train observers thoroughly and extensive fieldwork), are still manageable exercises.

However, measuring only the behavior component of performance may impede advances made in program impact evaluation. Limited behavior-related measurements are ultimately unsatisfactory if they cannot be linked to meaningful results-oriented frameworks, which require development assistance agencies to demonstrate that effective FP/RH programs are those that not only achieve high levels of quality, but also produce higher levels of access and use of services.

Accomplishments as Performance

Some HPT practitioners from business and industry have advocated the sole use of accomplishments as a summary measure of performance. Nickols, for example, states that performance is defined "by the outcomes of behavior" (2000). The rationale is simple: If workers' behavior and the quality of their product is substandard, that will be reflected in low sales, return on investment or some other indicator of accomplishment. By concentrating on the outcomes, proponents of this approach argue that they are simplifying measurements and are getting to the "bottom line." So if services are not used, or if sales falter, they will analyze the factors (including behavior) that may be affecting desired outcomes. In this scenario, behavior is treated as an intermediate variable of less value to business than the revenue product of the sales.

However, measuring accomplishments might not be adequate or even feasible in the context of FP/RH programs. First, information systems that track worker' accomplishments are not as readily available in the public and private health sector in developing countries as they are in business or industry. In FP/RH program evaluations, pre/post or time series designs have used client records and other facility-based statistics to obtain data for the accomplishment component of performance. These measurements obviously have relied on the quality of data from such records, which is of varying standards in many developing countries. There are also significant costs associated with the development of such information systems or sources in RH/FP programs in developing countries, and this may in part explain the reliance on behavioral measures.

Second, because RH/FP programs are human or social services, it seems necessary to focus attention on the processes involved in the delivery of services as well as the consequences or results of those services (i.e., measuring the quality of care in terms of technical competence, interpersonal skills or information given to clients, in addition to increased availability of services). Indeed, it is precisely because of the realization that simply demographic or even health indicators could no longer measure successful RH/FP programs, that clients' perspectives and rights, QoC and other similar paradigms have been developed. These have profoundly affected the way in which programs are designed and implemented nowadays.

Third, given the difficulty of measuring accomplishments alone, the additional measurement of provider behavior lends credibility to performance measurement results. Measuring both allows us to document links between provider behavior and service results, including its sequential nature. There has been some controversy about the influence of quality of care on contraceptive continuation (Cleland and Ali, 1994). Recent studies have demonstrated links between quality of care and outcomes such as use-continuation of services (Mensch et al., 1996; Koenig et al., 1997; Mroz et al., 1999; Ali, 2001). However, quality elements and methodologies have varied widely in such studies and other factors have also been found to affect client outcomes. This reinforces the argument for the need to maintain both the behavior and the accomplishment aspects in the definition.

As mentioned earlier, an additional difficulty in the use of service delivery outcomes as a measure of performance is determining what these outcomes represent. In almost all developing country settings, registration of client attendance and care at clinics is not linked to a specific provider. At the primary health care levels where one person is the sole provider of services, measuring provider outputs may be a reasonable task. However, the question "who is responsible for what?" inevitably arises when a number of providers are responsible for offering services at one SDP. A measure of "proportionate care" can be obtained through careful observations in clinics, deriving proportional attribution of client outcomes. Unfortunately, this methodology is only theoretical and might be difficult to put in practice for regular program evaluations.

Compounding these previous provider-related challenges is another common constraint regarding attribution of outcomes: the increasing concomitance of interventions. For example, interventions outside the scope of a training project, such as updating and disseminating guidelines or improving other personnel support structures and systems can influence provider behavior in unknown ways. IEC campaigns or health reform initiatives (e.g. free maternal health services), may also affect client access and use of services in varying degrees. Here, "proportional" attribution might prove an impossible task. Such aspects need to be considered in evaluation designs. For example, control groups in an experimental or quasi-experimental design might sort out the relative advantage of a particular intervention compared to "other things being equal." Multivariate analysis techniques may also elucidate relative contributions of each factor.

Future Directions

Despite the difficulties detailed above, neither *behavior* nor *accomplishments* used separately or together are likely to provide a complete picture of provider *performance* in RH/FP programs. It seems advisable, then, to maintain and find ways to operationalize the two-component HPT definition of performance. Our analysis has concentrated on both components of *performance* and advocates the use of both for a variety of reasons. In this section, we consider ways to operationalize their joint measurement.

At a recent meeting on the subject of performance measurement (see Acknowledgments), after reviewing each element through evaluations conducted by several international development assistance organizations, an attempt was made to arrive at a consensus regarding a common definition of performance and useful measurement methodology. One suggestion was the combination of separate behavior and accomplishments measurements into a single indicator. This might be done first by summarizing measurements of behavior in an index for subsequent use. Multi-item tools (e.g. checklists)

have been created to facilitate the measurement of quality of care (e.g. MEASURE, 2001) and composite indices of quality of care have already been developed and tested (Mensch et al., 1996; Koenig et al., 1997). A summative index could be related to a desired standard, producing a *relative* measure of behavior. Similarly, a summative figure could be obtained from percentage increases in client outcomes due to interventions, and compared to planned or desired objectives. The relative "effectiveness" (i.e. achieved/desired) found within each component could also be combined to arrive at a single indicator of average "performance".[1] This approach, while simple and straightforward, poses some questions around index construction (e.g. whether components should have equal weights), data level interpretation (e.g. mixing intermediate with outcome measurements) and the validity of assumptions (e.g. attribution; see above). There is a clear need for further research on the conceptual and methodological dimensions of using a single indicator to represent performance.

Conclusion

Human performance is not solely a measurement of the quality with which a job is done, nor is it only the results of the way a job is carried out—it is both. In the words of Nicholas (2001), "performance is the actual output and quality of work performed by organizations, teams or individuals." Despite the absence of tested summary indicators of performance, we propose that FP/RH program officers and evaluators resist the temptation to use the term *performance* to refer to one or the other component of the total of performance. We recommend that the term "performance" be used in reports and dissemination materials that describe evaluations that have looked at both behavior and accomplishments. In such way we will achieve greater conceptual clarity and sounder measurements.

Acknowledgments

The authors wish to thank Catherine Elkins of the MEASURE project for her valuable comments on an earlier draft of this paper, and to Marc Luoma, Lucy Harber, Nancy Kiplinger, Cathy Murphy, and David Nelson for their review of subsequent drafts. Thanks to Ellen Clancy and Kristin Immermann for editorial/bibliographic review. Some of the material used in this paper derives from "Measuring Provider Performance: A Technical Meeting" co-sponsored between PRIME II/Intrah and the MEASURE/*Evaluation* project, held December 6th and 7th, 2001 at the University of North Carolina, Chapel Hill. The authors are also grateful to the presenters for their contributions. More information about the meeting is forthcoming.

References

Ali, M. 2001. "Quality of Care and Contraceptive Pill Discontinuation in Rural Egypt." *Journal of Biosocial Science*. 33: 161-172.

Ashton, J. 2001. "Using performance improvement to increase vaccination rates in the Guatemalan Highlands." University Research Corporation, presentation at session 6021.2, *Performance Improvement – a Common Model Focusing on Health Results*, 129th Annual Meeting and Exposition, APHA, October 21-25, Atlanta, GA.

[1] The actual way of combining separate measurements is in debate. At the meeting a simple –unweighted average was presented. However, a suggestion was made (León, 2001) to multiply both components rather than adding them, to reflect the imperative need to include and improve both components. If one is missing the product would be "zero" performance. However, this might prove too stringent a criterion.

Bessinger, R. 2001. "Quick Investigation of Quality (QIQ): Assessing provider performance". Presentation at *Measuring Provider Performance: A Technical Meeting*. PRIME II/Intrah and MEASURE, The University of North Carolina at Chapel Hill, December 6th and 7th.

Binder, C. 2001. "Measurement: A Few Important Ideas." *Performance Improvement* 40 (3):20-28.

Brethower, D. 2001. "Healthcare Performance and the Bottom Line." Presentation at *Measuring Provider Performance: A Technical Meeting*. PRIME II/Intrah and MEASURE, The University of North Carolina at Chapel Hill, December 6th and 7th.

Broad, M. L. (ed.).1997. "Transferring learning to the workplace." Alexandria, VA: American Society for Training and Development.

Carlyle, J. J. and T. Ellison. 1984. "Appendix B: Developing Performance Standards" in Bernardin, John and Richard Beatty. *Performance Appraisal: Assessing Human Behavior at Work*. Boston, MA: Kent Publishing Company: 343-347.

Cleland, J, and M. Ali. 1994. "Quality of Care and Contraceptive Continuation," in Gao Ersheng and Iqbal Shah (Eds.), *Progress of Social Science Research on Reproductive Health*, Anthology of Treatises of the International Symposium on Social Science Research in Reproductive Health. Shanghai, People's Republic of China, 11-14 October. China Population Publishing House.

Dean, P. (Ed.). 1994. *Performance engineering at work.* Batavia, IL: International Board of Standard for Training, Performance and Instruction.

Dean, P. and D. Ripley. 1997. *Performance Improvement Pathfinders: Models for Organizational Learning*. Washington, D.C.: The International Society for Performance Improvement.

De la Peza, L. 2001. "Lessons from the field: Strengthening supervision using a performance improvement approach," Management Sciences for Health, presentation at session 6021.2, *Performance Improvement – a Common Model Focusing on Health Results*, 129th Annual Meeting and Exposition, APHA, October 21-25, Atlanta, GA.

Fronczak, N. 2001. "Service Provision Assessments – SPAs." Presentation at *Measuring Provider Performance: A Technical Meeting*. PRIME II/Intrah and MEASURE, The University of North Carolina at Chapel Hill, December 6th and 7th.

Gilbert, T. F. 1978. *Human Competence: Engineering worthy performance.* New York, NY: McGraw-Hill Book Company.

Harless, J. 1992. *The Accomplishment-based Curriculum Development (ABCD) System*, Harless Performance Guild.

Huntington, D., C. Lettenmaier, and I. Obeng-Quaidoo. 1990. "User's Perspective of Counseling Training in Ghana: The "Mystery Client" Trial." *Studies in Family Planning*, 21(3):171-177.

JHPIEGO. 2001. *Performance Improvement*. Data from http://www.jhpiego.org/global/pi.htm. Accessed February 2002.

Kaufman, R. 1997. *Preparing Performance Indicators and Objectives*, in Kaufman, Roger, Thigarajan, Sivisailam, and MacGillis, Paula, (eds), *The Guidebook of Performance Improvement: Working with Individuals and Organizations*. San Francisco, CA: Jossey-Bass Publishers.

Kirkpatrick, D. L. 1998. *Evaluating Training Programs: The Four Levels."* San Francisco, CA: Berrett-Koehler Publishers.

Koenig, M., M. B. Hossain, and M. Whittaker. 1997. "The Influence of Quality of Care upon Contraceptive Use in Rural Bangladesh." *Studies in Family Planning*, 28(4): 278-289.

Lindsley, O. 1999. "*Performance is Easy to Monitor and Hard to Measure."* ISPI proceedings article. Lawrence, KS: Behavior Research Company.

Luoma, M., W. Jaskiewicz, J. McCaffery, and D. Catotti. 2000. *PRIME Technical Report 19: Dominican Republic Performance Improvement Project Evaluation*. Chapel Hill, NC: Intrah.

McCaffery, J., M. Luoma, C. Newman, S. Rudy, A. Fort, and F. Rosensweig. 1999. *Performance Improvement Stages, Steps and Tools*. Chapel Hill, NC: Intrah.

MEASURE *Evaluation*. 2001. "Quick Investigation of Quality (QIQ) A User's Guide for Monitoring Quality of Care in Family Planning." *MEASURE Evaluation Manual Series, No. 2*. Chapel Hill, NC: Carolina Population Center, University of North Carolina at Chapel Hill.

Mensch, B., M. Arends-Kuenning, and A. Jain. 1996. "The Impact of the Quality of Family Planning Services on Contraceptive Use in Peru." *Studies in Family Planning,* 27(2):59-75.

Mielke, E., K. Beattie, A. Kaniauskene, and D. Adrience. 2001. "Using COPE® and performance improvement for community involvement." EngenderHealth, presentation at session 6021.2, *Performance Improvement – a Common Model Focusing on Health Results*, 129th Annual Meeting and Exposition, APHA, October 21-25, Atlanta, GA.

Mohrman, A. M. Jr, S. Resnick-West, and E. Lawler III. 1989. *Designing Performance Appraisal Systems*. San Francisco, CA: Jossey-Bass Publishers.

Mroz, T., K. Bollen, I. Speizer, and D. Mancini. 1999. "Quality, Accessibility, and Contraceptive Use in Rural Tanzania." *Demography* 36(1): 23-40.

Newman, C. 2002. "Following Up Performance: Lessons from The Field." *Performance Improvement*, 41(1): 11-18.

Nicholas, D. 2001. "Standards, Quality Improvement and Performance Measurement". Presentation at *Measuring Provider Performance: A Technical Meeting*. PRIME II/Intrah and MEASURE, The University of North Carolina at Chapel Hill, December 6th and 7th.

Nickols, F. 2000. *Performance and Performance Standards: Opinion*. Data from http://home.att.net/~nickils/opinion.htm. Accessed February 2002.

Population Council. 1998. *Reproductive Health Operations Research*. INOPAL III. New York, NY.

PRIME. 1997. *Regional Performance Evaluation* Workshop. Luknow, India. May 7-15. Unpublished.

Robinson, D. G. and J. C. Robinson.1990. *Training for Impact*. San Francisco, CA: Jossey-Bass Publishers.

Stillman, P. and M. A. Gillers. 1986. "Clinical Performance Evaluation in Medicine and Law" in Berk, Ronald, *Performance Assessment: Methods and Applications*. Baltimore, MD: The Johns Hopkins University Press.