

Scientific Program Evaluation for Health Development

Beverly HENRY¹⁾, PhD, Hon DSc.
Reiko UEDA¹⁾, D.M.Sci.

Abstract

Program evaluation provides direction for future scientific programs and new innovations in existing programs. Scientific program evaluation at four levels such as (1) evaluating reaction, satisfaction (2) evaluation learning (3) evaluating behavior (4) evaluating results was proposed. Then, authors concluded as following.

Program evaluation provides direction for future programs and new innovations in existing programs. An aim is to provide findings about a program which can be generalized to other contexts and periods of time. Achieving a degree of generalizability requires scientific research design and methodologically sound instrumentation. In nursing, program evaluation should bridge the gap between nursing education and nursing services. Effective teachers and nurses in hospitals and communities work together cooperatively to encourage learning, behaviors, and results -- with new strands of opinion -- that meet health needs and solve people's health problems. Having valid data about nurses' competence, their performance, and the effect of these on the nursing services is essential for sound health development. Scientific program evaluation contributes greatly to making sound decisions about programs and their contribution. Program evaluation involves rethinking everything from the program's learning model and curriculum, to the production of professional nurses and their proper performance in the health services for health development.

Key words:(1)program evaluation (2)health service (3)setting program objectives
(4)evaluation at 4 levels (5)nursing education and service

Toyama, former Minister of Education, Cultures, Sports, and Technology (MEXT) noted that, until about a decade ago, Japan lacked a climate conducive to university evaluation¹⁾. Then in the 1990s, self-evaluation was introduced for universities to improve the quality of their programs. In 1997 the University Council on Education Reform asked that self-evaluation to improve program performance be introduced into higher education using qualitative and quantitative research designs²⁾. Arima (2002), former president of Tokyo University, challenged those in higher education asking that universities pay more attention to basic and applied research: Program evaluation is an applied research methodology³⁾. As a university professor in the United States, Dr. Arima admired America's rigorous evaluation of programs.

The purpose of program evaluation is to continuously develop the quality of a program's performance - *kieretsu*. Theory-based program evaluation can be designed in many ways, but all designs for scientific evaluation should begin during a program's planning phase⁴⁾. Planning an effective program starts with multitasking to determine learning needs, then entails setting program objectives, determining learning content for implementation, coordinating, and evaluating program results.

In any quality improvement effort - and program evaluation is performed for quality improvement - a useful design includes addressing internal and external factors at four levels⁵⁾. The two internal factors are: 1) student reaction and satisfaction and 2) mastery of program content as measured through learning. The external factors are: 3) behavioral change as demonstrated by workplace performance, and 4) the results of behavioral changes in terms of their effect on the quality of health services.

Program Evaluation at Four Levels

1. Evaluating Reaction, Satisfaction

Firstly, evaluating reaction and satisfaction is done for short-term assessment of the quality of the subject material taught and the skill of instructors. Evaluating reaction using paper and pencil tools is the easiest.

1) Okinawa Prefectural College of Nursing

Therefore, not surprising, it is the most widely used type of evaluation. Data are also collected about reactions and satisfaction with facilities, schedules, and instructional aids. In reaction surveys, open-ended items may be added to closed-items to encourage qualitative comments. Reaction sheets are completed before the close of a program and returned immediately. Acceptable numerical standards are decided on ahead of time by evaluators who often use scaled numerical values such as excellent=5 to poor=1. Statistical tests can be calculated (e.g. paired comparisons) to strengthen validity when analyzing random or non-random groups.

2. Evaluating Learning

Secondly, participant learning is evaluated. The questions are what knowledge has been gained? What skills have been developed? How have attitudes improved? Behavior cannot change until learning takes place. Student learning is a measure of instructor effectiveness. To evaluate learning designs, use a control group that does not receive the program intervention. The control and experimental groups should be as similar as possible. Cost and availability determine random or nonrandom assignment throughout a program's evaluation. Evaluate learning before and after a program unless the content is entirely new, in which case only post-testing is possible. Use paper and pencil tests to measure knowledge, use performance tests to measure skills, and ensure a 100% response rate. Perform item analysis on tests and conduct further tests to indicate the impact of the program over a period of time.

3. Evaluating Behavior

Thirdly, evaluating behavior in the workplace - worker performance - is more difficult and time-consuming than evaluating reaction and learning. Use a control group when possible. Focus groups may also be helpful. A key question at this level, with implications for validity is: When to evaluate behavior? A sufficient amount of time must pass for behavioral change in performance to take place. Two to three months is a good general rule. Measurement of pre- and post-program change is usually triangulated through surveys and interviews by supervisors, and subordinates, or those best qualified to observe the visual clues of the desired newly learned behaviors. Patterned interviews are best in which all interviewees are asked the same questions about participants' performance. Costs and benefits of behavioral evaluation should also be factors in obtaining a 100% response. Remember, positive results cannot occur if behavior does not change. Several caveats: Behavior may change, but results may stay the same. Moreover, changes to some new behaviors may result in unanticipated negative results. And some behaviors encouraged in the classroom may not be appropriate for all participants at work in the health services.

4. Evaluating Results

Fourthly, evaluating results is the final and most difficult level. It is also referred to as summative evaluation inasmuch as results summarize a program's accomplishments. Whenever possible, experimental designs and longitudinal analyses should be employed⁶. The evaluator's main concerns are with improved quality, productivity, and safety. For nursing education, questions like the following are asked. What are the tangible benefits of the program compared to its costs? How have health indicators in the community improved? How has the program affected interdisciplinary teamwork and interpersonal cooperation in hospitals? If practical, use a control and allow the proper amount of time to pass for results to be measurable using tests, questionnaires, and observations. Repeat measurements as appropriate. Evaluating results determines whether a program is continued. Benchmarking for comparison is especially useful at this stage to spot unnecessary costs and eliminate redundancies⁷. At each stage of program evaluation, there are many designs and valid instruments that can be adapted for nursing from published studies. Program evaluation studies may be high or low-profile initiatives depending on their costs and goals.

Conclusions

Program evaluation provides direction for future programs and new innovations in existing programs. An aim is to provide findings about a program which can be generalized to other contexts and periods of

time. Achieving a degree of generalizability requires scientific research design and methodologically sound instrumentation. In nursing, program evaluation should bridge the gap between nursing education and nursing services. Effective teachers and nurses in hospitals and communities work together cooperatively to encourage learning, behaviors, and results -- with new strands of opinion -- that meet health needs and solve people's health problems. Having valid data about nurses' competence, their performance, and the effect of these on the nursing services is essential for sound health development. Scientific program evaluation contributes greatly to making sound decisions about programs and their contribution. Program evaluation involves re-thinking everything from the program's learning model and curriculum, to the production of professional nurses and their proper performance in the health services for health development.

References

- 1) Toyama, A.: University reform in Japan to usher in a century of knowledge. Tokyo, Japan. JSPS, Japanese Society of the Promotion of Science, Quarterly. 9; 4, 2004.
- 2) Higa, Y.: The role of Japanese university in the twenty-first century. Tokyo; Japan Council for Education Reform, 1997.
- 3) Arima, A.: The future of higher education in Japan. Third Annual Michio Nagai Memorial Lecture. Tokyo: United Nations University. 2002.
- 4) Fitzpatrick, JL, Sanders, JR, & Worthen, BR.: Program evaluation. 3rd Edition. Boston: Pearson, 2004.
- 5) Kirkpatrick, DL.: Evaluating training programs. The four levels. San Francisco, Berrett-Koehler, 1998.
- 6) Taylor Fitz-Gibbons, C. & Morris, LL.: How to design a program evaluation. Newbury Park, California: Sage, 1987.
- 7) Henry, BM.: Handbook of practical research for quality improvement in health care: A new approach to nursing research. Tokyo, Igaku Shoin, 2004.

Suggested Reading

McKenzie, JF. & Smeltzer, JL.: Planning, implementing, and evaluating health promotion programs. Boston, Allyn and Bacon, 3rd Edition, 2001.

抄録

科学的プログラム評価

—ヘルスディベロップメントのために—

ビバリー ヘンリー¹⁾ 上田 礼子¹⁾

<抄録>

これからの科学的プログラム評価の方向および現在あるプログラムの刷新について述べた。科学的プログラム評価の4つのレベルとして①反応、満足の評価、②学習の評価、③行動の評価、④結果の評価を提示した。その上で、著者らは次のような結論に達した。

看護のプログラム評価は看護教育と看護サービスの溝に橋を架けるものである。看護師のコンピテンシー、パフォーマンス、そして健康サービスの有効性へのそれらの影響に関する有用なデータを所持することは高度に機能する健康システムの基本である。プログラム評価の研究は個人の目標や費用によって位置付けに高低があるかもしれない。有能な教員と病院・地域の看護職者が、新しい意見をもち、人々の健康上のニーズに出会い、健康問題を解決する学習、行動、結果を促進するために協力し合って共に働くことである。科学的プログラム評価は、プログラムについての健全な意思決定および地域的、国家的、国際的に高度に機能するヘルスシステムへのプログラム創出に大きく貢献するものである。

<キーワード>

①プログラム評価 ②健康サービス ③目標設定 ④評価の4段階 ⑤看護教育とサービス

1) 沖縄県立看護大学