

Measuring success in the global village - resources for conducting systematic and comprehensive evaluations in IT settings

Gail Marshall

Gail Marshall & Associates

2393 Broadmont Ct., Chesterfield, MO 63017, USA

Tele + 314 230 6613; Fax, + 314 230 3609

E-mail: 74055.652@Compuserve.com

Abstract

A well-designed evaluation plan is essential if decision-makers are to assess the problems and prospects for Information Technology. In designing the evaluation plan the resources in the community, the point of view guiding teaching and learning, and lessons from previous attempts at innovation must all be taken into account. Many models for conducting evaluations exist. Tests, measures and alternative assessment strategies currently in circulation can be adapted for IT settings. The dissatisfaction with attempts to use IT successfully in education, voiced by many commentators around the world, could be countered with data on the powerful impact IT on schools if all the relevant data were collected, analyzed and used in successive cycles of policy making, program improvement and program evaluation.

Keywords

Assessment/testing, case studies, evaluation, policy, culture

1 INTRODUCTION

Every society has been 'technological'. Across Africa, for example, many different technologies supported life. The inhabitants of the Olduvai Gorge had spatial techniques for mental and physical map making and direction finding. They

developed implements suitable for a wide variety of animal and plant husbandry. Gold mining and cattle farming contributed to the economy long before machines and microchips began to mesmerize societies around the world. Villages have traditional ways of assessing how well they are operating and we should be careful not to abandon village ways that provide valuable information as we try to collect data on the impact of Information Technology (IT) on schools.

Developing countries, now beginning to implement IT in many schools and seeking to assess IT's impact, should be cautious about adopting the pre/posttest model of evaluation to the exclusion of other data collection techniques. Adopted as part of the 'efficiency' movement in American education, the pre/posttest model, if used at all, should only be used as one data collection strategy near the end point of an evaluation cycle. At the beginning of the cycle many more strategies must be employed - strategies that will describe and analyze the match between what was anticipated, what was accomplished and what went wrong. Every village storyteller knows how to capture the essence of a situation and so, in the global village, we should be relying on a wide range of skills and strategies to convey the essence of successful endeavours and point out where we went wrong.

Both early and recent IT adopters face a variety of implementation, assessment and institutionalization challenges - among them the question, in the case of early adopters, of whether current national, local and classroom IT policies are appropriate; in the case of recent adopters the question is whether other countries' IT policies and practices will meet their own needs. In establishing national policies and local projects both groups of countries must consider the importance of early, frequent and targeted evaluations of IT planning, implementation and impact. Resources, used in evaluating a wide range of pre-IT innovations, can be adopted and adapted to provide answers to all the questions that must be asked before we ask, 'Did it make a difference for the students?'

2 NECESSARY AND SUFFICIENT CONDITIONS FOR EVALUATING IT'S IMPACT

2.1 The need for systematic and comprehensive evaluations

Evaluations of IT in educational settings must be systematic and comprehensive. Successful IT use changes the dynamic at all levels of use (Rowe, 1996); unsuccessful use introduces dislocations and disaffections across the organization. Evaluations must look at a wide range of activities and events - the planning process, the initiation process and the implementation process - and evaluations must occur for several cycles before any pretest/posttest comparisons between student groups occur. The evaluator must be a full partner in the innovative effort. A wide variety of measures must be used and several different types of reports must be prepared - each addressing specific facets of the innovative process.

2.2 Defining goals for IT use

Very early in the life of the innovation it is important to ask, 'What are our goals and how will we know we have achieved them?' Answering such a question is not

easy because each participant in the innovative process may have, especially at the early stage of the project, only partially-formed ideas about what should be done and how success will be defined. The problem of differing ideas held by members of the innovation group has been especially true with many technology-based innovations where the content and processes are often so novel that many decision-makers are unaware of or do not understand the implications of curricular choices. Database and telecommunications use, for example, have often been poorly understood, even by educators well-versed in other aspects of technology. So the introduction or evaluation of these aspects of technology has been problematic.

An important, although often lengthy, part of the development of an evaluation plan is the negotiation of what will be evaluated and how it will be evaluated (Smorodin, Marshall and Brooks, 1981). Such negotiations are often ignored or curtailed but at the risk of finding, at the end of the evaluation cycle, that each participant had a different vision of IT and that few, if any, of the participants' real questions have been answered.

2.3 Recognizing different epistemological viewpoints and different educational objectives

Another neglected dimension, the epistemological point of view, must be discussed before, during and after evaluation activities begin. At the present time the question of which point of view should determine the design and delivery of instruction is hotly debated (Marshall, 1995). For example, the introduction of Integrated Learning Systems (ILSs) in the United States, the United Kingdom and Australia has been welcomed by some educators and resisted by others. A major reason for the debate is the irreconcilable difference between educators with a behaviorist view of learning and educators with a constructivist view of learning (Reese and Overton, 1970). The worst-case scenario for an evaluation is that teaching and learning based on constructivist principles is evaluated within a behaviorist framework - and only that framework - or vice versa. Important questions that flow from one theory will be ignored while other questions, which may not be germane, are asked instead. Benzie (1995) stresses the importance of understanding how the questions we ask about IT use determine policy for good or ill.

Different historical foundations for the design and delivery of instruction also mean different designs of instruction, which call for different measures of success. For example, a school or a country may choose from among one or more of the following foundations of instruction - each with its own goals and objectives: (1) learning in the classical tradition, which emphasizes the Western canon of literature, history and science; (2) learning as a socialization process, which emphasizes the historical forms of interactions within a village or community; (3) learning as a vocational activity, which mimics the actions and goals of industry and commerce; (4) learning as a political activity, which seeks to inculcate the ideas and values of the ruling (or, in some cases, the wishing-to-be-in-power disenfranchised) class.

IT, while it appears to be value neutral, can be applied to any one of these goals for learning. Evaluating IT's impact on students can only be accomplished when the evaluators decide which point of view is actually being espoused at the school

and/or classroom level. Lengthy sessions may be needed until all parties agree on the central themes and the interesting questions related to the themes. Failure to spend time at this critical point may lead to failed evaluations as implementation progresses since the evaluator's questions may focus on aspects of the project ('Have students' keyboarding skills improved?') that conflict with others' visions ('How much more fluent have students become in writing via computer?').

3 RESOURCES FOR EVALUATING IT

3.1 National policies, individual strengths and cultural traditions

It is perfectly acceptable if one of the greatest resources for planning, implementing and evaluating a country's IT use is its own history of innovation and implementation. Strategies that have worked for other innovative efforts may lend themselves to the IT endeavour. Peoples' experiences, both positive and negative, in attempting other change processes are invaluable assets for supporting the IT revolution. Similarly, the country's culture and traditions must be carefully considered when IT policies are established. For example, while factions in the highly industrialized nations may call for classroom-based IT use to support large doses of drill and practice because the tradition of the 'school as factory' is widely established as cultural currency, other factions may cite pedagogical and cultural traditions that dictate very different ways of using IT resources.

Resources from other countries and from international organizations, in terms of policy statements and practices for determining effectiveness, may provide models that should be considered as evaluations of IT are framed. Many countries' professional bodies - the National Council of Teachers of Mathematics in the USA and the Association of Mathematics Teachers in UK, for example - are also valuable resources. The publications of the International Federation of Information Processing (IFIP) may also facilitate the evaluation of IT use.

It is especially important for countries to learn from one another, both with respect to failures as well as successes. For example, the rationalist model of policy-making assumes that a carefully planned series of decisions will lead to a coherent set of practices but the reality of many innovative efforts prior to the IT revolution shows us that change is not straightforward, and is often compromised by individual and collective activities that deform and derail plans (House, 1975). Policy-makers in different countries should analyze their own efforts against the yardstick of other countries, asking how what works in one country is transferable to other countries and how problems in some situations might prove instructive for the design of future policies and practices.

People who have played an active role in all prior educational activities and innovations are significant resources for evaluating IT's impact. Administrators, teachers, students, parents and other community members have long memories. They will remember previous attempts at innovation. They will be able to offer their analyses of what went wrong and what went right. At this and all other stages let tradition play its proper role. In some communities storytelling is a hallowed and powerful way of transmitting data. Data from stories told about innovative

failures and successes may be far more useful than data collected via surveys and questionnaires.

Throughout the innovative process there are at least two contexts in which the innovation is being conducted and judged. On the one hand there are national and local education authority goals deciding school inputs and processes. Those national and local bodies also set criteria for judging outputs; on the other hand the family, the students and the community-at-large have goals, and provide their own inputs (student motivation to work with the innovation or parents' satisfaction or dissatisfaction with the goals and processes, for example). A well-designed evaluation will use community resources to assess the impact of IT implementation at all levels of the country and community.

3.2 Models for evaluations

When it comes to evaluating IT use in schools we seem to have regressed to a primitive stage of evaluation. Far too many evaluation designs are conducted in a small number of classrooms using a pretest/posttest experimental design. As Rowe (1996) points out, we should be looking for differences in patterns - how teachers in IT settings do what they do differently and how those differences bring about differences in students' learning - and not patterns of differences. Many sophisticated but uncomplicated techniques have been used in pre-IT days and should be studied as models. For example, *Federal Programs Supporting Educational Change Volumes I - VIII* (Berman and Pauly, 1975) describes a complex design for evaluating four federally funded change projects in the USA.

The Rand researchers conducted a nationwide survey of 293 sites - including interviews with project participants at all levels of responsibility. Then 29 field studies - at sites with different characteristics and goals - were conducted with observations, interviews and analyses of the change process as a major part of the evaluation effort. Volume I of the Rand Corporation's report contains a lengthy bibliography of change-oriented efforts prior to 1974.

The Eight Year Study, an analysis of attempts to modernize the curriculum in selected American high schools during the 1940s, remains a classic of its kind. Tests, inventories, questionnaires, checklists and student logs in 30 high schools were used to see if planned changes in curriculum actually occurred (Smith & Tyler, 1942). More recent analyses of curriculum innovation can be found in Fullan and Pomfret (1977).

The case study model provides crucial data for deciding how the early stages of IT implementation are proceeding. The case study should be an important way to ensure that plans are proceeding on schedule, that goals have been specified, and that problems are being identified and addressed. As the implementation process begins, the evaluator observes the decisions and actions of all participants, and interviews them frequently. Interviews are especially useful during the staff development phase and early stages of classroom implementation. Cooley and Lohnes (1976) have published examples of several case studies, including an analysis of guidance counseling programs, an evaluation of bilingualism and a study of primary education in Ireland. Such models will be useful for IT-based evaluators.

A variant on the case study is the Action Research approach to collecting data. Teachers and others work together to discuss and analyze patterns of teaching and learning, looking for generalizable trends. However one does not embark on an Action Research approach without taking the time to fully acclimate all participants its methods and without recognizing the demands for data collection that it places on teachers (Jaworksi, 1994).

For many policy-makers and community members the experimental design model is the only one considered appropriate for evaluating the impact of an innovation. It is regarded as 'scientific' and 'value free' - both designations assumed to legitimize the use of the experimental design above all other designs. In fact, the model is often misused. It is often used to the exclusion of other more necessary data collection activities and the results are often interpreted inappropriately. Other techniques and evaluation models are far more appropriate. For example, Worthen and Sanders (1973) have collected many different models of evaluation, including Provus's discrepancy model and Stufflebeam's call for evaluations as decision-making tools.

Evaluation should be a long-term endeavour. Way before the classic pretest/posttest design is used all sorts of data should be collected. It is important to know that staff development occurred and was deemed satisfactory; that classroom instruction occurred and was managed in such a way that students could be seen to be using IT as well as applying skills and concepts gained from IT use. Only when all questions about implementation have been answered is it appropriate to ask, 'Does using IT make a difference?'

3.3 Models for reporting

The design and intent of the interim and final reports should be one of the first items on the evaluator's agenda. S/he must discuss with the rest of the innovation group the goal of the information to be collected through the evaluation. Shipman (1989) cites Bastiani's discussion of models (1978) as a useful way to frame reporting. The models are:

- the 'Basic Information Model' where only factual information is transmitted - the number of computers purchased, the student-to-computer ratio, etc.;
- the 'Public Relations Model' where persuasion or the attempt to avoid conflict by describing what has been accomplished is the goal;
- the 'E.P.A. Primary Model' where an attempt is made to convince one constituency of the value of the innovation while trying to convince another constituency of the need for resources;
- the 'Academic Recruitment Model' which seeks to recruit adherents or participants - either recruiting teachers to participate or urging parents to support the program by enrolling their children;
- the 'Developmental Model' where information is provided to support participants as they go through the early stages of implementation;
- the 'Involvement Model' where the information is used to create a deeper involvement of all participants.

In fact, each of these models may be used at key points in the process of initiating and sustaining IT use.

3.4 Instruments for observation and analysis

There are time-honored methods of collecting data, including storytelling. But at times other methods are called for, including the use of easy-to-scan-and-score instruments. Instead of devising instruments, the evaluation team should consider using instruments that have served other innovators, either by using the instruments as designed or modifying them to fit the situation. *Mirrors for Behavior* (1974) is a compendium of classroom observation instruments ranging from the Adams-Biddle instrument that enables educators to analyze the sociology of a classroom to the Wright-Proctor instrument designed to assess the extent to which teachers were using new pedagogical methods in 'new math' classrooms.

It is often that case that the most successful implementations can be undone by choosing inappropriate measures. For example, if students work steadily at problem-solving software in their mathematics classes and are then tested with an instrument that merely measures basic computational ability we should not be surprised that the test results show no difference between computer-using and noncomputer-using students. Most basic computation tests are too limited to assess the range of problem solving skills acquired through the use of problem-solving software.

Fortunately several resources are available for choosing well-designed and well-targeted tests. The Buros *Mental Measurements Yearbook* series lists tests in subject areas ranging from achievement tests in English and other languages to tests that measure clerical and mechanical ability. Descriptions of the tests, analyses of their psychometric properties, and information on how to order and administer the tests can be found in the Buros collections.

Once again, professional organizations around the world have developed curriculum frameworks and assessment strategies that lend themselves to IT settings. The State of California, for example, has a wide range of publications devoted to assessment. Stenmark's (1989) presentation of alternative assessments is only one example.

4 CONCLUSION

Data provide the power for decision makers to continue or refine policies and practices. In the absence of data implementation may falter, aspirations may wither and dreams for the future may die. It is essential that technology use at all levels, from the national to the student level, be assessed in order to determine problems, generate solutions and drive further IT development. Carefully assembled, the wealth of evaluation resources can highlight difficulties and identify successes.

5 REFERENCES

Benzie, D. (1995). The Impact of Our Questions on Information Technology Policies and Practice, in *Proceedings of the World Conference on Computers*

- in *Education VI: Liberating the Learner* (eds. J.D. Tinsley and T.J. van Weert), Chapman & Hall, London.
- Berman, P. & Pauly, E. (1975) *Federal Programs Supporting Educational Change*. Rand Corporation, Santa Monica, CA.
- Buros, O. *Mental Measurements Yearbooks*. Buros Institute, Lincoln, NE.
- Cooley, W.W. and Lohnes, P.R. (1976) *Evaluation Research in Education*. John Wiley and Sons, New York.
- Fullan, M. and Pomfret, A. (1977) Research on curriculum and instruction implementation. *Educational Researcher*, 47(2).
- House, E. (1975) *The Politics of Educational Innovation*. McCutchan, Berkeley, CA.
- Jaworski, B. (1994) *Investigating Mathematics Teaching: A Constructivist Enquiry*. Falmer Press, London.
- Marshall, G. (1995) The ABC's of today's technology decision-making in Australia, the United Kingdom and the United States. A paper presented at the WCCE '95 World Conference Computers in Education, Birmingham, UK.
- Reese, H.W. and Overton, W.F. (1970). Models of development and theories of development. In *Lifespan Developmental Psychology: Research and Theory* (eds. L. R. Goulet and P. B. Baltes), Academic Press, New York.
- Rowe, H. (1996) I.T. is failing to revolutionise the curriculum, because to date we have failed to evaluate its benefits in context. Proceedings the Australian Computer Society, Canberra, Australia.
- Shipman, M. (1989) *Assessment in Primary and Middle Schools*. Routledge, London.
- Simon, A. and Boyer, E.G. (1974) *Mirrors for Behavior III: An Anthology of Observation Instruments*. Research for Better Schools, Philadelphia, PA.
- Smith, E.R. and Tyler, R.W. (1942) *Appraising and Recording Student Progress*. Harper and Row, New York.
- Smorodin, C., Marshall, G. and Brooks, T. (1981) Point of View: A Critical Variable in Program Development and Program Evaluation. A paper presented at the American Educational Research Association meeting, Los Angeles, CA.
- Stenmark, J. (1989) *Assessment Alternatives in Mathematics: An Overview of Assessment Techniques that Promote Learning*. Regents, University of California, Berkeley, CA.
- Worthen, B.R. and Sanders, J.R. (1973) *Educational Evaluation: Theory and Practice*. Charles A. Jones Publishing Co., Worthington, Ohio.

6 BIOGRAPHY

Gail Marshall evaluates projects for local, state and federal agencies. She was a Divisional Assistant in the Division of Research and Evaluation for the St. Louis Public Schools, a member of the evaluation team for the Comprehensive School Math Project (CSMP) and the Midwest Regional Exchange at the Central Midwest Regional Laboratory She was also the coordinator of the IBM Model Schools Project 1983-1984 for the St. Louis Public Schools. A graduate of Marymount Manhattan College, New York, she has an MA and a Ph.D. from Washington University, St. Louis, where she studied child development and psychometrics.

A member of IFIP's WG 3.5, Gail Marshall is also a member of the Computer Using Educators, based in California, USA, the International Society for Technology in Education (ISTE) and she has been listed in *Who's Who in American Education*. The international editor for ISTE and the software review editor for *Education Technology Monitor*, Gail raises irises and many different varieties of artemesia.