

# THE DANCE OF EVALUATION: HUSTLE OR MINUET

---

Eva L. Baker

*Center for the Study of Evaluation  
UCLA Graduate School of Education*

The practice of education is poised on an insecure technological base. Despite massive infusions of rhetoric, little progress has been made to transform the operations of instruction into other than blurred attempts at craft. At present, the scientific aspect of education is limited principally to discipline based research conducted with far-term outcomes. In

only restricted areas, under highly controlled conditions, have the responsibilities of the education profession derived benefit from scientific analysis.

Dance is an art form with a technological base and a transpositional motive. Although dance and educational practice are both performed *en chorale*, they display more particular similarities. Consider the references to the hustle and the minuet, two well-known dances. While separated greatly in historical time, they

both share common elements. They, for instance, were both popular expressive forms, based on highly controlled progressions of rigidly stylized steps. Thus, their suitability as metaphors for evaluation activities should be immediately perceived. As evaluations, the hustle and minuet differ along other continua, from energetic to passive, or exuberant to reserved. Notice, however, that these dances, as evaluation, both contribute at a marginal level to the serious pursuits of their times.

Emerging technology in the field of education mirrors none of the art but some of the other attributes of these dances. For example, the tendency in education is to codify and stylize operations, e.g., technological procedures, and grind at them insistently either until some higher authority, usually political, questions ones' motives, or until a new procedure is proffered. Some procedures are adopted with high hopes and energy and others fit more clearly into the established order. In instructional development, specifically, we have seen almost mechanical commitment to different procedures and guidelines: rules for creating, first, linear programs, then, branching instruction, and then, multi-component programs, have been articulated. The field of evaluation, at once older and less mature than that of instructional development, seems destined to continue the same errors. For evaluation, as it became regularized into models, frameworks, and procedures, has spawned a series of rules for action which rarely receive sustained reflection. A review of the role of evaluation in instructional planning and development is suggested, with a focus on an examination of the purpose and utility of evaluation for educational technology.

A critical question, almost lost in the clutter of contending models and procedures of evaluation, must be raised about who is to assume the evaluation role. Functions generate labels and labels have a way of becoming anthropomorphized. Technology created first technicians, then technologists. Development transmuted into developer; evaluation became evaluator. This linguistic fact is of minor interest except for the connotation such designators suggest. Who is best qualified to evaluate a development project? An evaluator or anyone else? Certainly, any right-minded person would choose the individual who is *named* with the function, for special expertise is clearly implied. Another problem with anthropomorphic process is that a complex set of activities, which requires participants with diverse preparation and expertise, is reduced to a simple procedure, evaluation, where a simple, all-purpose individual, the *evaluator*, can handle the entire show.

Should, then, development give over the evaluation function to evaluators? If development activity is stripped of its evaluation component, what is left? A pessimistic view of the state of applicable findings from instructional research would suggest that, without evaluation,

very little work remains for those charged with development responsibility. They might create instructional design based on a few stable principles of learning. They may base program structuring decisions on rather firm but far between findings from cognitive psychology. Development personnel may make media selections by rehearsing any one of a number of sets of arbitrary, and usually, economically inspired rules. They may manage component development and integration by using systems such as PERT, Critical Path Analysis or other planning procedures. But the relinquishment of evaluation functions to another external group gives up the most potent development capability there is, the capacity to detect and manage change. It is evaluation, at least at present, that drives the development act.

Development can be conceived as a continuum where we begin in ambiguity and attempt to move toward clarity. At the outset of a development project, we may only have the sketchiest information about what we will end up with. For example, we will know what age or experience learners we are designing materials or curriculum for... probably. We will know the broad area of subject matter we are to treat... probably. We will know how much we have to spend... definitely. A few overzealous people might even have outcomes to which the curriculum or product is directed specified in advance. During the course of the development effort, we attempt to gain better understanding of our process and what our likely outcomes will be. We learn to identify learners with precise areas of need and history for our products. We determine what settings will form the probable contexts of product use. We design, revise, and redesign component tests of programs in order to formulate a proper effort which embodies appropriate instructional characteristics. In the end, we hope that we can say that the product we have developed is planned for a particular group or groups of respondents for use in a specified class of settings and with results reliable at given levels, when appropriate time and implementation guidelines are met. It should be clear that throughout this process, the need to evaluate is pervasive.

Instructional development staff may have a wide range of expertise in the evaluation area, but they usually share some familiarity with the concepts of formative and summative evaluation. As described by Scriven (1967) and

Markle (1967) using slightly different terms, formative and summative evaluation are assumed to have different purposes. Summative evaluation should provide a comparative judgment of merit under conditions as objective as possible. The purpose of this form of evaluation is to make a selection among competing alternatives, a topic we will return to later.

Formative evaluation differs in its purpose for data are collected to improve a program under development, to allow adjustments to be made in order to maximize the desired outcomes of the program. Thus, the formative evaluation contributes to the redevelopment and refinement of the product as it shifts along the continuum from ambiguity to precision. It is clear that instructional development must keep the reins on formative evaluation activity.

Rather than guidelines, some simple rules of thumb can be described for formative evaluation. First, one should collect extensive information from relatively few subjects during initial design phases. This suggestion is based on three probable states of the world. First, the early versions of the product and program may be poor. Why expose great numbers of subjects to unreliable instruction? Second, it is likely that significant aspects of the program are susceptible to revision, for instance, the objectives, formats, instructional approaches and so on. Thus, diverse information sources are desirable. Third, the practical matter of processing data from many subjects is a serious impediment. Development momentum can be stopped and the fluidity of the project lost by extensive data analyses and interpretation requirements at the wrong time in a project's life. Another major guideline is that the diversity of data one collects reduces as the project moves toward completion. Essentially one collects less information, but from more different kinds of people. It should be remembered that conservation of resources requires that only usable information be assembled, that is, data which can be employed to revise the program. The need to be comprehensive should be overcome in an effort to make the evaluation act instrumental to the efficient completion of the project.

With regard to the relationship of summative evaluation, that is the program verification, it is probably not in the development staff's or developing agency's interests to conduct a summative or comparative evaluation. Whenever what is known as "summative evaluation" seems

appropriate in a development cycle, the decision for product selection will likely be surrounded by political as well as effectiveness criteria. For this reason, development personnel might seriously consider avoiding the entire prospect of summative evaluation. A brief reflection of the actual conditions of summative or comparative evaluations might make the point more strongly.

First, data from any comparison between curriculum or products are likely to be inclusive. This likelihood, in fact, is the basis of our parametric statistics. Thus, there will be a force for the *status quo*, and a tendency against innovation when the decision is based purely on empirical data. Second, in a comparison between a commonly formatted program, e.g., printed materials, in contrast with multimedia development, the short run utilitarian aspects of the contending programs, when divided by costs, will almost always support the more conservative development effort. The long range benefits, for instance, of introducing stimulating variation in school activities, are, in the first place, almost never assessed by appropriate dependent measures, and in the second place, the amount of time appropriate for such a long term comparison is usually impractical for all participants. In addition, in-

novative looking materials may be threatening both to the users of instructional programs, teachers, and administrators, and to the funders of education, for example, state legislators.

The paradox is that while development people recognize the importance of comparative product tests in their activity, the summative evaluation has little to recommend itself to product development staff on a strictly practical basis.

Further, unless summative evaluations are contracted externally to the organization of the developed product, they are inevitably suspect. Summative evaluations should be conducted independently by agencies or groups for whom there are no contingencies for supporting the effectiveness of the product under study. (Who really believes the Pepsi challenge advertising campaign, conducted, as it is, by Pepsi-Cola employees. Perhaps if Canada Dry Ginger Ale monitored the comparisons, we would have developed a greater sense of trust in the findings.)

If summative type evaluations are to be conducted by the development agency, however, they should be limited to a few clear purposes. For instance, they seem to be appropriate for promotional materials required for broad implementa-

tion. Data from summative evaluations can also be used formatively, perhaps to improve user manuals. But summative evaluation asks potentially to document our failures without recourse. Thus, it must be the minuet of this piece, a contrived exercise of limited practicality for development staff. Formative evaluation is where the hustle, energy and productivity intersect and continue to contribute to the improvement of instruction. For current evaluation of our government-supported and for-profit efforts as well as for the persisting longevity of instructional development itself, we should subscribe more evaluation models derived more from Motown than from Mozart.

#### References

- Markle, S. M. Empirical testing of programs. In P. C. Lange (ed.) *Programmed Instruction*, 66th Yearbook of the National Society for the Study of Education, Part II. Chicago: NSSE, 1967. pp. 104-138.
- Scriven, M. The methods of evaluation. In R. W. Tyler, R. M. Gagné, and M. Scriven (eds.) *Perspectives of Curriculum Evaluation*, AERA Monograph Series on Curriculum Evaluation, No. 1, Chicago: Rand McNally, 1967. pp. 39-83.