1. PROGRAM EVALUATION: A HISTORICAL OVERVIEW

GEORGE F. MADAUS and DANIEL L. STUFFLEBEAM

Program evaluation is often mistakenly viewed as a recent phenomenon. Many people date its beginning from the late 1960s with the infusion by the federal government of large sums of money into a wide range of human service programs, including education. However, program evaluation has an interesting history that predates by at least 150 years the explosion of evaluation during the era of President Johnson’s Great Society and the emergence of evaluation as a maturing profession since the sixties. A definitive history of program evaluation has yet to be written and in the space available to us we can do little more than offer a modest outline, broad brush strokes of the landscape that constitutes that history. It is important that people interested in the conceptualization of evaluation are aware of the field’s roots and origins. Such an awareness of the history of program evaluation should lead to a better understanding of how and why this field has developed as it did.

Where to begin? For convenience we shall describe seven periods in the life of program evaluation. The first is the period prior to 1900, which we call the Age of Reform; the second, from 1900 until 1930, we call the Age of Efficiency and Testing; the third, from 1930 to 1945, may be called the Tylerian Age; the fourth, from 1946 to about 1957, we call the Age of Innocence; the fifth, from 1958 to 1972, is the Age of Development; the sixth, from 1973 to 1983, the Age of Professionalization; and finally the seventh from 1983 to 2000 the Age of Expansion and Integration.
THE AGE OF REFORM 1792–1900

We begin this period in our history of program evaluation in 1792 because that is
the year in which William Farish invented the quantitative mark to score examina-
tions (Hoskins, 1968). Replacing qualitative assessments of student performance with
a mark for a “correct” answer permitted the ranking of examinees and the averag-
ing and aggregating of scores. This was the first development in the field of
psychometrics as we know it today (Madaus & Kellaghan, 1992). In fact Farish
revolutionized testing, a technology that plays an important role in the history of
program evaluation to the present.

The 19th century was the era of the Industrial Revolution with all of its atten-
dant economic and technological changes. The very structure of society was trans-
formed. Major social changes occurred. There was drastic change in physical and
mental health and outlook, in social life and social conscience, and in the structures
of social agencies. There was the laissez-faire philosophy of Bentham and the
humanitarian philosophy of the philanthropists (Thompson, 1950). There were
continued but often drawn out attempts to reform educational and social programs
and agencies in both Great Britain and the United States.

In Great Britain there were continuing attempts to reform education, the
poor laws, hospitals, orphanages, and public health. Evaluations of these social
agencies and functions were informal and impressionistic in nature. Often they
took the form of government-appointed commissions set up to investigate aspects
of the area under consideration. For example, the Royal Commission of Inquiry
into Primary Education in Ireland under the Earl of Powis, after receiving testi-
mony and examining evidence, lamented over the progress of the children in the
national schools of Ireland. The Powis Commission recommended the adoption of
a scheme known as payment by results, already being used in England, whereby
teachers’ salaries would be dependent in part on the results of annual examinations
in reading, spelling, writing, and arithmetic (Kellaghan & Madaus, 1982; Madaus &
Kellaghan, 1992). Another example of this approach to evaluation was the 1882
Royal Commission on Small Pox and Fever Hospitals, which recommended after
study that infectious-disease hospitals ought to be open and free to all citizens
(Pinker, 1971).

Royal commissions are still used today in Great Britain to evaluate areas of
concern. Rough counterparts in the United States to these commissions are presi-
dential commissions (for example, the President’s Commission on School Finance),
White House panels (e.g., the White House Panel on Non Public Education), and
congressional hearings. Throughout their history royal commissions, presidential
commissions, and congressional hearings have served as a means of evaluating human
services programs of various kinds through the examination of evidence either gath-
ered by the Commission or presented to it in testimony by concerned parties.
However, this approach to evaluation was often only emblematic or symbolic. N. J.
Crisp (1982) captures the pseudo nature of such evaluations in a work of fiction.
One of his characters discusses a royal commission this way: “Appoint it, feel that