Introduction

Thomas Kellaghan
Educational Research Centre, St. Patrick's College, Dublin, Ireland

Daniel L. Stufflebeam
The Evaluation Center, Western Michigan University, MI, USA

Lori A. Wingate
The Evaluation Center, Western Michigan University, MI, USA

Educational evaluation encompasses a wide array of activities, including student assessment, measurement, testing, program evaluation, school personnel evaluation, school accreditation, and curriculum evaluation. It occurs at all levels of education systems, from the individual student evaluations carried out by classroom teachers, to evaluations of schools and districts, to district-wide program evaluations, to national assessments, to cross-national comparisons of student achievement. As in any area of scholarship and practice, the field is constantly evolving, as a result of advances in theory, methodology, and technology; increasing globalization; emerging needs and pressures; and cross-fertilization from other disciplines.

   The beginning of a new century would seem an appropriate time to provide a portrait of the current state of the theory and practice of educational evaluation across the globe. It is the purpose of this handbook to attempt to do this, to sketch the international landscape of educational evaluation – its conceptualizations, practice, methodology, and background, and the functions it serves. The book's 43 chapters, grouped in 10 sections, provide detailed accounts of major components of the educational evaluation enterprise. Together, they provide a panoramic view of an evolving field.

   Contributing authors from Africa, Australia, Europe, North America, and Latin America demonstrate the importance of the social and political contexts in which evaluation occurs. (Our efforts to obtain a contribution from Asia were unsuccessful.) Although the perspectives presented are to a large extent representative of the general field of evaluation, they are related specifically to education. Evaluation in education provides a context that is of universal interest and importance across the globe; further, as history of the evaluation field shows, the lessons from it are instructive for evaluation work across the disciplines. In

fact, many advances in evaluation stemmed from the pioneering efforts of educational evaluators in the 1960s and 1970s.

Contemporary educational evaluation is rooted in student assessment and measurement. The distinction between measurement and evaluation, suggested by Ralph Tyler more than 50 years ago and later elaborated on by others, had an enormous influence on the development of evaluation as an integral part of the teaching and learning process. For many years, educational evaluation focused mainly on students' achievements; it concentrated on the use of tests and was immensely influenced by psychometrics. Another major and long-standing influence on educational evaluation is to be found in a variety of programs to accredit schools and colleges. Mainly a U.S. enterprise, accreditation programs began in the late 1800s and are an established reality throughout the U.S. today.

It was only in the mid-1960s and early 1970s, with the increased demand for program evaluation made necessary by various evaluation requirements placed on educational programs and projects by governmental organizations and other agencies, that educational evaluation dramatically expanded and changed in character. While earlier evaluation, as noted above, had focused on student testing and the educational inputs of interest to accrediting organizations, the new thrust began to look at a variety of outcomes, alternative program designs, and the adequacy of operations. To meet new requirements for evaluation, evaluators mainly used their expertise in measurement and psychometrics, though they also took advantage of two other resources: research methodology and administration. Research methodology – mainly quantitative but later also qualitative – provided the guidance for data collection procedures and research designs that could be applied in evaluation. Administration theory and research helped to improve understanding of planning and decision making, which evaluations were designed to service, as well as of the politics of schools.

Most developments in program evaluation took place in the United States and were "exported" to other parts of the world, sometimes only ten or twenty years later. In Europe, for instance, the major concern was – and in some countries still is – testing and student assessment, although tests and other achievement measures have begun to be used for other purposes. Gradually, tests came to be used as outcome measures for other evaluation objects, such as programs, schools, and education systems, sometimes alongside other information regarding the objects' goals and processes. Widely varying applications of evaluation can now be found around the world in many shapes and sizes, reflecting its far-reaching and transdisciplinary nature.

Side by side with all this activity, evaluation has been growing into a fully fledged profession with national and international conferences, journals, and professional associations. It is practiced around the world by professional evaluators in universities, research institutes, government departments, schools, and industry. It is being used to assess programs and services in a variety of areas, such as criminal justice, welfare, health, social work, and education. Each area, while having much in common with evaluation in general, also has its unique features.

Three distinctive features set educational evaluation apart from other types of evaluation. First, it has been strongly shaped by its roots in testing and student assessment, on one hand, and curriculum and program evaluation on the other. In other areas (e.g., health services or criminal justice), evaluation focuses mainly on programs and is usually considered as a type of applied research. Although it took many years for educational evaluation to come to the point where it would not be perceived only as student assessment, such assessment is still an important element of the activity. Second, education is the predominant social service in most societies. Unlike business and industry, or other social services such as health and welfare, education affects, or aspires to affect, almost every member of society. Thus, public involvement and the concerns of evaluation audiences and stakeholders are of special significance in educational evaluation, compared to evaluation in other social services, and even more so when compared to evaluation in business and industry. Third, teachers play very important roles in educational evaluation as evaluators, as evaluation objects, and as stakeholders. They are a unique and extremely large and powerful professional group, with a high stake in evaluation and a long history as practicing evaluators assessing the achievements of their students, and must be taken into account whenever evaluation is being considered.

Education is one of the main pillars of the evaluation field, and thus it is important that those who work in educational evaluation should be part of the general evaluation community, participating in its scientific meetings and publishing their work in its professional journals. There is much that they can share with, and learn from, evaluators in all areas of social service, industry, and business. However, educational evaluators should also be sensitive to the unique features of their own particular area of evaluation and work to develop its capabilities so that they can better serve the needs of education and its constituents. It is our hope that this handbook will aid members of the educational evaluation community in this endeavor.

The handbook is divided into two parts, Perspectives and Practice, each of which is further divided into five sections. While the individual chapters can stand on their own as reference works on a wide array of topics, grouping them under Perspectives and Practice, provides in-depth treatments of related topics within an overall architecture for the evaluation field. In the first part, the perspectives of evaluation are presented in five major domains: theory, methodology, utilization, profession, and the social context in which evaluations are carried out. The second part of the handbook presents and discusses practice in relation to five typical objects of evaluation: students, personnel, programs/projects, schools, and education systems. Chapters in the handbook represent multiple perspectives and practices from around the world. The history of educational evaluation is reviewed, and the unique features that set it apart from other types of evaluation are outlined. Since the chapters in each section are ably introduced by section editors, we will only comment briefly on each section's contents.

The opening section deals with perspectives on educational evaluation by examining its theoretical underpinnings. Ernest House introduces the section by

noting that scholars have made substantial progress in developing evaluation theory, but remain far apart in their views of what constitutes sound evaluation. Michael Scriven provides an overview and analysis of theoretical persuasions, which may be grouped and contrasted as objectivist and relativist. Specific evaluation theory perspectives presented in the section include Daniel Stufflebeam's CIPP model, with its decision/accountability and objectivist orientations; Robert Stake's responsive evaluation, that stresses the importance of context and pluralism and advocates a relativist orientation; the constructivist evaluation approach of Egon Guba and Yvonna Lincoln, with its emphasis on participatory process and rejection of objective reality; and the relatively new entry of democratic deliberative evaluation, advanced by Ernest House and Kenneth Howe, which integrates evaluation within a democratic process. Other sections present additional theoretical perspectives, including ones relating to utilization-focused evaluation, participatory evaluation, connoisseurship evaluation, and experimental design. Readers interested in the theory of evaluation will find in these chapters ample material to support dialectical examination of the conceptual, hypothetical, and pragmatic guiding principles of educational evaluation.

Section 2 focuses on evaluation methods. Evaluators, as the section editor Richard Wolf notes, differ in their methodological approaches as much as they differ in their theoretical approaches. Major differences are reflected in the extent to which investigators control and manipulate what is being evaluated. At one extreme, randomized comparative experiments, described by Robert Boruch, are favoured; at the other extreme, completely uncontrolled, naturalistic studies, described by Linda Mabry. Other methods presented in the section include cost-effectiveness analysis, described by Henry Levin and Patrick McEwan, and Elliot Eisner's educational connoisseurship approach. In general, the section reflects the current dominant view that evaluators should employ multiple methods.

The chapters in Section 3 provide in-depth analyses of how evaluators can ensure that their findings will be used. Section editor Marvin Alkin and his colleague Carolyn Huie Hofstetter summarize and examine research on the utilization of evaluation findings. Michael Patton and Bradley Cousins, respectively, present state-of-the-art descriptions of utilization-focused and participatory models of evaluation, and explain how they foster the use of findings.

Section 4 editor Midge Smith acknowledges that the evaluation field has made progress toward professionalization, yet judges that the effort is still immature and in need of much further thought and serious development. The topics treated in the section include Daniel Stufflebeam's report on progress in setting professional standards, Michael Morris's treatise on evaluator ethics, Blaine Worthen's examination of the pros and cons of evaluator certification, Lois-ellin Datta's analysis of the reciprocal influences of government and evaluation, Hallie Preskill's proposal that the evaluation field become a sustainable learning community, and Midge Smith's projection of, and commentary about, the future of evaluation. Overall, contributors to the section characterize evaluation as an emergent profession that has developed significantly but still has far to go.

Section 5 editor Harbans Bhola notes that the practice of evaluation is, and should be, heavily influenced by the local social setting in which the evaluation is carried out, but also characterizes a global context for evaluation. Particular settings for educational evaluation are discussed in chapters by Carl Candoli and Daniel Stufflebeam for the U.S., Ove Karlsson for Europe, Fernando Reimers for Latin America, and Michael Omolewa and Thomas Kellaghan for Africa. Contributions to the section make clear that evaluation practices are heavily influenced by a nation's resources and employment of technology, as well as by local customs, traditions, laws, mores, and ideologies. A clear implication is that national groups need to set their own standards for evaluation.

Section 6 editors Marguerite Clarke and George Madaus introduce chapters on the assessment of student achievement, which has been, and continues to be, a core part of educational evaluation. This is the kind of assessment that impacts most directly on students, often determining how well they learn in the classroom or decisions about graduation and future educational and life chances. It takes many forms. Robert Mislevy, Mark Wilson, Kadriye Ercikan, and Naomi Chudowsky present a highly substantive state-of-the-art report on psychometric principles underlying standardized testing. Peter Airasian and Lisa Abrams describe classroom evaluation practice, which arguably is the form of evaluation which has the greatest impact on the quality of student learning. Caroline Gipps and Gordon Stobart describe concepts and procedures of assessment that have received great attention in recent years in response to dissatisfaction with traditional methods, particularly standardized tests. Thomas Kellaghan and George Madaus provide a description of external (public) examinations and issues that arise in their use in a chapter that 20 years ago would probably have evoked little more than academic interest in the United States. However, having long eschewed the use of public examinations, which have a tradition going back thousands of years in China and form an integral part of education systems today in many parts of the world, the United States over the last decade has accorded a form of these examinations a central role in its standards-based reforms.

Section 7 editor Daniel Stufflebeam argues that educational evaluations must include valid and reliable assessments of teachers and other educators, and that much improvement is needed in this critical area of personnel evaluation. According to Mari Pearlman and Richard Tannenbaum, practices of school-based teacher evaluation have remained poor since 1996 but external programs for assessing teachers, such as teacher induction and national certification assess-ments, have progressed substantially. According to Naftaly Glasman and Ronald Heck, the evaluation of principals has also remained poor, and shows little sign of improvement. James Stronge ends the section on an optimistic note in his report of the progress that he and his colleagues have made in providing new models and procedures for evaluating educational support personnel. Overall, the section reinforces the message that educational personnel evaluation is a critically important yet deficient part of the educational evaluation enterprise.

James Sanders, the editor of Section 8, drew together authors from diverse national perspectives to address the area of program/project evaluation.

Program evaluation as practiced in developing countries is described by Gila Garaway; in the U.S.A. by Jean King; in Canada by Alice Dignard; and in Australia by John Owen.

Section 9 editor Gary Miron introduces the chapters on evaluation in schools with a discussion of the old and new challenges in this area. Daniel Stufflebeam offers strategies for designing and institutionalizing evaluation systems in schools and school districts; James Sanders and Jane Davidson present a model for school evaluation; while Robert Johnson draws on his work with Richard Jaeger to provide models and exemplars for school profiles and school report cards. Catherine Awsumb Nelson, Jennifer Post, and William Bickel present a framework for assessing the institutionalization of technology in schools.

While Section 6 deals with the evaluation of individual students, the chapters in Section 10 address the use of aggregated student data to evaluate the performance of whole systems of education (or clearly identified parts of them) in a national, state, or international context. As section editor Thomas Kellaghan points out, the use of this form of evaluation grew rapidly throughout the world in the 1990s as education systems shifted their focus when evaluating their quality from a consideration of inputs to one of outcomes. Two major, and contrasting, forms of national assessment are described in the section. Lyle Jones describes the sample-based National Assessment of Educational Progress in the United States, while Harry Torrance describes the census-based assessment of the national curriculum in England. William Webster, Ted Almaguer and Tim Orsak describe state and school district evaluation in the U.S. Following that, international studies of educational achievement, the results of which have been used on several occasions to raise concern about the state of American education, are described by Tjeerd Plomp, Sarah Howie, and Barry McGaw. William Schmidt and Richard Houang write about a particular aspect of international studies, cross-national curriculum evaluation.

We hope that the handbook will be useful to readers in a variety of ways, helping them to consider alternative views and approaches to evaluation, to think about the role and influence of evaluation in national settings, to gain a better understanding of the complexities of personnel and program evaluation, to gain perspective on how to get evaluation findings used, to look to the future and to the steps that will be needed if evaluation is to mature as a profession, to identify a wide range of resource people, to appreciate the needs for evaluation at different levels, and to identify common themes to ensure integrity for evaluation across national settings.

We are indebted to the authors, section editors, and others who contributed to the handbook. David Nevo assisted the publisher in starting the Handbook project, contributed to the overall plan, and helped select and recruit some of the section editors and chapter authors. We acknowledge the assistance of Michael Williams of Kluwer Academic Publishers, who consistently supported our effort. Hilary Walshe, Regina Klöditz, and Rochelle Henderson at the Educational Research Centre at St Patrick's College in Ireland and Sally Veeder at the Evaluation Center at Western Michigan University provided competent editorial, technical, and clerical assistance.