

TEACHERS AS EVALUATORS PROJECT

POLICY PAPER

A Rationale for Curriculum Evaluation in Australia

A DRAFT FOR COMMENT

Not to be quoted or reprinted

Prepared by: Phillip Hughes
Neil Russell
Diana McConachy

A RATIONALE FOR CURRICULUM EVALUATION IN AUSTRALIA

I. BACKGROUND TO THE TEACHERS AS EVALUATORS PROJECT.

The project arose as a result of discussions taking place over more than a year and involving Canberra College of Advanced Education (C.C.A.E.), Curriculum Development Centre (C.D.C.), Educational Research and Development Committee (E.R.D.C.), the Commonwealth Department of Education, Schools Commission and State Departments of Education.

Although there was considerable concern and debate about the quality of education, there was remarkably little information available for the people most centrally concerned to make judgments about current needs and the means for improvement. Further, with the growing emphasis on the school as the base for curriculum development it became clear that there was a need for the project to focus on those processes and activities by which a school could improve its own curriculum.

The initiation of the project in 1978 coincided with the publication of the Report of the Curriculum Evaluation Study Group¹, sponsored by C.D.C., a report which gave support to the general thrust of the project and provided much useful base material.

The initial discussions led to the formation of the project, with the following objectives:

- a. To develop an understanding of, and commitment to, the role of evaluation processes as means of curriculum improvement.
- b. To involve and train teachers in evaluation processes directed towards curriculum improvement.
- c. To develop materials designed to help teachers towards this end.
- d. To involve curriculum officers and inservice personnel at the system level in the evaluation process directed towards:
 - (i) assisting teachers develop their own evaluation processes
 - (ii) evaluating their own procedures and programs.
- e. To develop materials designed to assist curriculum officers and inservice personnel working with teachers on school-based evaluation.
- f. To develop new patterns for the continued training of teachers.

The thrust of all these emphases is directed towards the improvement of educational programs. A great deal of information is available relating to performance in schools: external and internal examinations, standardised test results, school and classroom tests, national projects such as those developed by A.C.E.R.² looking at particular areas, and international projects such as those sponsored by the UNESCO Institute of Hamburg, the I.E.A. Studies³. There is, in addition, a considerable amount of information in the form of reports by inspectors and education advisers and case studies by research workers. Very little of the wider range of information available has been used by or made available to those who are centrally involved in curriculum improvement, educators in schools or service activities. With the growing emphasis on the school as the locus for curriculum development, the need has developed for a project to work with teachers and others centrally involved on the use of currently available information for curriculum decisions and on the development of other necessary information. This is the basis of the Teachers as Evaluators Project, T.E.P.

II. PATTERNS IN THE UNITED KINGDOM AND THE U.S.A.

The debate on education in countries with similar cultures such as Australia, Canada, the United Kingdom and the United States exhibits common features. Three types of questions seem to occur commonly in public debate:

- . what is the place of formal assessment in education for student placement or selection, e.g. external examinations for certification and tertiary entry?
- . how can education systems/schools inform various audiences in appropriate fashion concerning the worth of their programs?
- . how can schools and systems develop procedures to improve the value and relevance of their own programs?

Each of these questions raises the use of evaluation in a particular sense:

1. For vocational or educational placement.
2. For accountability.
3. As curriculum evaluation.

Our concern here is mainly with the third aspect of public debate, i.e. curriculum evaluation but with a realisation of the possible inter-connections of all three modes.

Australian education has drawn freely from two major sources of information with respect to both ideas and practice - the United States and the United Kingdom. This general pattern applies to

the important area of curriculum development and to the newly emerging and related emphasis in curriculum evaluation.

301

In the United States the major source of methodology in educational evaluation lies in the various models derived from psychology. For the past decade a number of writers have pointed to the limitations of such models in educational evaluation, e.g. Glass⁴, while other writers such as Hastings⁵ have pointed to the potential contribution of methods from several disciplines.

The main pattern of curriculum evaluation for some time still derived from the writings of Ralph Tyler⁶ and the development of the objectives model of curriculum design. The model identified four major elements:

- purposes, goals, aims, objectives.
- learning activities.
- teaching methods.
- evaluation.

The model has been subject to two major forms of change:

- a. The refinement of the meaning of objectives as an aspect of the statement of purposes.
- b. An elaboration of the model itself to express the essentially cyclic and dynamic nature of curriculum development.

The best-known name with respect to the refinement of objectives is Benjamin Bloom who was associated with two major publications in the area, the first relating to cognitive objectives⁷ and the second to affective objectives⁸. These publications sought to develop a classification of objectives in terms of observable student behaviour, a classification which would permit a systematic analysis of the whole spectrum of desired outcomes in an unambiguous and meaningful way.

The Bloom classification has been criticised by a number of writers but still remains one of the most frequently used forms of analysis of educational outcomes. The criticism has been on points such as the following:

1. Pre-specifying the educational objectives limits the planner and teacher to those objectives predicted in advance and prevents the utilisation of unexpected but desirable outcomes.
2. Since it is easier to specify objectives in areas such as distinct psycho-motor skills or the rote learning of information there is a danger that more complex objectives or those more difficult to define may be undervalued.

3. The model separates means and ends in a way which is unjustifiable in practice.

The argument is a continuing one. A major reason for the persistence of the objectives model in the face of such argument is that the argument tends to be directed against extreme forms of the model. Thus, with respect to Point 1 above, the objective approach does not hinder in any necessary way, the teacher from taking into account unexpected outcomes. Similarly, while it is easier to specify objectives in some area than others, Bloom's classification does provide useful check-lists against over-emphasis or under-emphasis. Finally, the model does not necessarily separate means and ends artificially. Particular aims of the teaching process can be also a part of the means of achievement, for example, taking part in discussions can be both a learning strategy and a desired aim.

A further major reason for the continued use of the objectives model has been the failure to develop an alternative pattern of comparable utility.

The development which has taken place in the U.S. and elsewhere is not so much the displacement of the objectives model as its subsumption into more inclusive patterns.

One need for extension of the pattern was pointed out by Cronbach⁹ in considering the uses of the evaluation. He pointed to three types of decision for which evaluation data may be used:

1. Course improvement: deciding on needed changes in instructional material and methods.
2. Individuals: decisions about individuals for planning instruction, for selection or placement, for informing student on progress.
3. Administration: judging effectiveness of the system, or of teachers, etc.

A further need for extension was pointed out by Stufflebeam¹⁰. Stufflebeam related evaluation to decision-making, as had Cronbach, but identified four decision types which led to the specification of four types of evaluation.

1. Context Evaluation: is directed to assist planning decisions in the determination of objectives. It analyses the situation and relates actual and desired conditions. Rather than accepting the given goal-statements it helps to provide a rationale for the determination of objectives.
2. Input Evaluation: relates to decisions on the design of the program and the use of resources to meet goals.
3. Process Evaluation: deals with implementation, identifying possible sources of failure and developing an interpretative description of events.

4. Product Evaluation: relates to the achievement of objectives and thus relates closely to the Tyler-Bloom model.

Stufflebeam's approach is usually described as the C.I.P.P. Model.

A further major development came with Michael Scriven's 'Goal-Free Evaluation.'¹¹ This approach deliberately kept consideration of goals by the evaluator out of the early stages of evaluation in order to allow the evaluator to view the total processes more objectively. Scriven's approach was implemented through his use of a product check-list which replaces consideration of the goals with an analysis of needs. The role of the check-point on the check-list is that failure to meet the requirements is an absolute disqualification for the product. His product check-list has thirteen check-points, each scored on a scale of 0-4, with scores of 0 or 1 defined as unacceptable.

The contribution of Robert Stake similarly was to broaden greatly the view of evaluation, for example, in his paper The Countenance of Educational Evaluation¹² where he stressed both the formal and the informal aspects of evaluation and the importance of evaluation as portrayal. Stake's model went on to describe projects for evaluation in terms of antecedents, transactions and outcomes and then to examine these in terms of both intention and the observed results. These patterns of development are mentioned here since they have their parallel in the United Kingdom and in fact there seems to be a remarkable degree of congruence in current thinking in spite of radically different starting points and educational contexts. In the United States, however, the evaluation movement is currently bogged down in a political battle at national and state levels concerning accountability.

In the United Kingdom external examinations are still firmly established through the examining boards and their position seems to be enhanced to the degree that their place is seldom questioned even by those who are most vocal in their comments on the unfortunate side effects of formal assessment. In fact current methods in this area seem to bring the much more flexible and informal C.S.E. (Certificate of Secondary Education) modes of examination under the same pattern as the Higher School Certificate examinations. It is strange that, in the arguments raging in the U.K. concerning both accountability and curriculum evaluation, the massive system of examinations is scarcely mentioned.

The second aspect of evaluation mentioned above, namely accountability, is a lively issue in the United Kingdom. Its most visible and controversial form is in the APU, the Assessment of Performance Unit. The Department of Education and Science set up the Assessment of Performance Unit in 1976 to carry out a comprehensive program to monitor the performance of children at school. The APU operates through three main types of group. First, there is a coordinating group consisting of teachers, education advisers, teacher educators and inspectors. This group recommends the areas of the curriculum to be examined and the working groups

needed to study how to monitor them. It subsequently coordinates and supervises the work. Secondly, there is a consultative committee set up by the Secretary of State for Education to examine the general priorities proposed for the work and to make recommendations with respect to them. This committee is representative of the main groups concerned with education - teachers, employing authorities, parents, employees and employers. Thirdly, there are working groups which also have representative patterns of membership and which are responsible for work within a specified area.

The areas defined are as follows:

1. Science
2. Mathematics
3. Language
4. Modern foreign language
5. Aesthetic development
6. Personal and social development
7. Physical development.

At the stage of writing only the first three working groups have developed published charters which are to be used as a basis of testing systems. The general patterns of policy, however, are clear. Assessment is to be based on a small number of key areas, each concerned with one important and distinct facet of the basic skills and attitudes that pupils develop. For example, the science group is looking at criteria under five headings - Observation, Generalisation, Explanation and Enquiry, Component Skills, Attitudes. The working groups are quite specifically not related to particular areas of the school curriculum, even in those cases where there are school subjects of similar names. For example, science is regarded as a term which describes particular ways of thinking about and tackling problems rather than as a particular school subject or group of subjects.

The terms of reference of the APU¹³ are:

"to promote the development of methods of assessing and monitoring the achievement of children at school and to seek to identify the incidence of under-achievement."

Its four tasks are:

- . to identify and appraise existing instruments and methods of assessment which may be relevant for these purposes
- . to sponsor the creation of new instruments and techniques for assessment having regard to statistical and sampling methods

- . to promote the conduct of assessment in cooperation with local education authorities and teachers

- . to identify significant differences of achievement related to the circumstances in which children learn, including the incidence of under-achievement, and to make the findings available to those concerned with resource allocation within the appropriate Department, local education authorities and schools.

It will be noted that the APU itself is designing its work in order that there cannot be comparisons made between particular schools and particular pupils. The methods of light sampling defined by the APU will ensure this. However, it is already apparent that a number of local education authorities see the tests developed by the APU as a means by which they can make deliberate comparisons. Certainly, the work of the APU is seen by many people working in evaluation in England as a very considerable threat to a more constructive use of evaluation. They see the APU, in spite of good intentions, as being used for distinctly political ends and the mention in the terms of reference of the use of information by those who distribute resources does little to allay this fear. Indeed the doubts are shared by many of those involved in the APU working groups. It is interesting to compare this situation to that which has arisen in the U.S. in states such as Michigan, and which is reported in the article by Jerome T. Murphy and David K. Cohen¹⁴. This shows the difficulties which arise from the uncritical use of assessment measures which make comparisons between schools and which are then proposed as the basis for the allocation of resources.

The rapid development of thinking in the area of curriculum evaluation is quite dramatically exhibited in two comparatively recent publications. The first is the book by Wiseman and Pidgeon, Curriculum Evaluation¹⁵. The second is the book by Tawney, David (ed) Curriculum Evaluation Today: Trends and Implications.¹⁶ The development in emphasis between the two books is striking as is the rise of a quite new generation of curriculum evaluators.

The book by Wiseman and Pidgeon builds quite specifically on the Tyler model, taking into account the developments by writers such as James Popham and Elliott Eisner¹⁷. Eisner introduced a distinction between "instructional objectives" (where outcomes can be defined with some precision) and "expressive objectives" (where the task is more open-ended, as in aesthetic appreciation). Wiseman and Pidgeon asserted that teachers had unconsciously avoided the Tyler approach because of a preference for implicit rather than explicit goals. They go on to recommend an approach based firmly on the definition of aims in precise terms. The book edited by Tawney, on the other hand, introduces a variety of additional approaches, arising from felt needs. These approaches echo the suggestions made by Hastings ten years earlier and establish evaluation procedures based on

historical, sociological and anthropological methods.

The source of this change can be quite clearly traced to the development of a large number of national curriculum projects, each of which required the appointment of an evaluator. Tawney lists thirty seven such projects, featuring prominently the Schools Council and, to a lesser degree, the Nuffield Foundation and the Departments of Education. The names currently associated with curriculum evaluation in England all tend to be of relatively young people and almost all arise from an assignment to a particular curriculum project. The prominent names include Wynne Harlen, Michael Eraut, Malcolm Parlett, David Hamilton, Barry Macdonald, John Elliott, Marten Shipman, Helen Simons, Denis Lawton, Paul Black, Peter Kelly, David Jenkins, and David Layton. It is certainly not correct to think of 'a group' of evaluators in the sense of the existence of a common approach. There is, however, an extension of the concept of evaluation and of the sources for evaluation models which bears a distinct similarity to recent U.S. thinking.

A key paper in the British evaluation scene was the publication by Parlett and Hamilton, Evaluation as Illumination¹⁸, first published in 1972 but reprinted in the book by Tawney. Their concern sprang initially from the comparative lack of success of many large curriculum projects both in the U.S. and the U.K. and their feeling that the classical model of evaluation was inadequate.

Parlett and Hamilton in their approach show distinct similarities to the concepts put forward by Stake,¹⁹ of evaluation as portrayal. The British authors draw a strong distinction between what they describe as the "agricultural paradigm" of evaluation, traditionally used, and their own preferred form, the 'social anthropology paradigm'. The approach depends heavily on participant observation and on the ethnographic field-work of social anthropology. Other evaluators since these have pointed out the possible relevance of literary criticism, investigative journalism and film documentary²⁰. Parlett and Hamilton use the term 'triangulation' to describe their three-stage approach: first, an overall study to identify significant features; second, the selection of a number of such features for more intensive inquiry; and third, the attempt at 'explanation' through seeking general principles underlying the organisation of such a program.

The Parlett-Hamilton paper was closely followed by a conference at Churchill College, Cambridge, in December, 1972 which, while not able to reach agreement of all issues, published a statement indicating some major shifts in concern. These were as follows:

- "1. that the traditional methods of evaluation had paid too little attention to the whole educational process in a particular milieu and too much attention to those

changes in a student behaviour which could be measured;

2. that the educational research climate had underestimated the gap between school problems and conventional research issues;
3. that curriculum evaluation should be responsive to the requirements of different audiences, illuminative of complex organisational processes, and relevant to both public and professional decisions about education."²¹

The work of Parlett and Hamilton on illuminative evaluation was developed further by those with an interest in case studies, including Barry Macdonald who was the evaluator in the Humanities Curriculum Project.²² In this project he made a quite deliberate and carefully considered move away from a psychometric approach to a style relying on methods usually associated with historical and anthropological research, for example, observation, interviewing and documentation. In addition, Macdonald re-defined the idea of audience. The consumers of his work were decision makers of various kinds, the sponsors (the Schools Council and Nuffield Foundation), the employing authority, the schools, and the examination boards.

As seen by Macdonald in the Humanities Curriculum Project, evaluation began to be seen in terms of supplying answers to questions that would be asked by the consumers or decision-makers. Even this was seen as too restrictive, since such a view assumed that questions could be specified in advance whereas it was felt that many questions would only arise during the process of the evaluation itself. This project therefore adopted a very ambitious combination of traditional and non-traditional methods, so ambitious in fact that it was never completely finished.

In the period since these publications a number of doubts have been expressed about the so-called sociological or anthropological paradigm:

1. While there are established rules of procedure for anthropologists working in unfamiliar societies it does not necessarily follow that they can be carried over into curriculum evaluation.
2. The rules of procedure for non-traditional evaluation are insufficiently clear and the skills need to be specified more clearly. The area possesses no tradition comparable to the established standards of historical and anthropological research.
3. The variety of possible interests and audiences places evaluators in a situation of role conflict which may pose considerable strain for the evaluator.

4. There is a danger of subjective impressions being put forward as objective data.
5. The methods and language developing in this field may become as esoteric and remote from teachers and other interested parties as has conventional educational research.

It is not intended that the above comments should dismiss this obviously important area but it does represent a series of problems which need to be solved.

The current situation in the United Kingdom seems in fact to have a number of unfortunate effects:

1. In spite of the intentions of those involved, there seems to have developed a separation between illuminative evaluation and traditional evaluation which is unnecessary and unfortunate.
2. There is also an artificial dichotomy between such areas as process evaluation and product evaluation even though this was not the original intention.
3. The debate over the "right of information" has frequently led to an over-emphasis on the right of information for one group of people, for example, teachers, with the implication being that teachers in a particular situation may have the right to refuse publication of information which shows their work in a critical light.
4. By the nature of the recent evaluation studies, the results are ephemeral. They are linked with particular curriculum projects and as soon as those projects cease to be funded they have generally ceased to operate and thus the information obtained cannot be related to current practice.
5. Similarly, because of the association of curriculum evaluation with particular projects there is a lack of direct and continuing links between curriculum evaluation theory and school practice.

There is thus a danger that the audience for evaluators may be seen as exclusively for the evaluators and the funders of projects. Perhaps the major danger to emerge is the tendency to think of evaluation in terms of clear-cut alternatives, for example, the classical model and the illuminative model. Many papers and discussions give the impression that the two styles of evaluation are separate and are leading to separate developments. This is clearly not the case. Non-traditional evaluation has evolved criteria out of certain shortcomings and dissatisfactions with

the experimental model and does not represent a complete alternative.

Blaine Worthen, in an interesting address to the 1977 Australian Association for Research in Education conference, discusses the problem of the evaluator faced by a variety of models.²³ Worthen, speaking from a wealth of practical experience as an evaluator as well as a wide knowledge of evaluation models, indicates that he had seldom, if ever, used a particular model as an entity in one of his own projects. "I couldn't think of a single lone-wolf evaluation of my own where I had consciously selected any single model to guide the study. Instead for several years I have been designing each evaluation *de novo*, pulling pieces of the models in as they seemed relevant. Certain features of some models I used frequently, others seldom or never." He goes on to argue that this is not so much a defect of current models as a reflection of a lack of a theoretical position with a sufficiently sound empirical base.

This viewpoint implies the need for a joint approach by those involved in evaluation, combining the careful use of current evaluation procedures in a variety of contexts and the examination of that use to research the assumptions and concepts involved.

III. IMPLICATIONS FOR PLANNING AND ACTIVITIES IN T.E.P.

There are some clear dangers existing in the current stage of thinking about curriculum evaluation which need to be avoided in the Australian Teachers as Evaluators Project.

One danger has already been mentioned. It is that the current lack of certainty and clarity in evaluation theory should lead to an attempt at a theoretical base. This could lead to a multitude of activities, without a real attempt to make explicit the underlying purposes and assumptions of the procedures, to examine the effects of their use and the role of the various participants.

Another danger is the reverse, that evaluation should become a specialized activity of the evaluators, with well-developed theoretical bases but little practical outlet. In this pattern, as in England, a few major projects could develop with a strong theoretical input but without the well-developed connection with teachers and schools to permit them to interact with practice.

To avoid these dangers the Teachers as Evaluators Project will require an orientation to both research and development. This approach requires a careful delineation of evaluation aims and procedures and the linking of these to particular evaluation

projects. These projects require clear understanding by all the participants to agreed roles. The continuing study of these situations should form the basis of an effort to derive some general principles.

In the time-scale available to the project, slightly over two more years, a clear definition of priorities is required if anything substantial is to be achieved. This definition will take into account what has emerged during the operation of the project during 1978 and will attempt to make clear the future priorities for policy and action.

A. CURRENT BASE FOR PROJECT

1. Steering Committee

This links T.E.P. with major education bodies in Australia, C.D.C., E.R.D.C., Schools Commission and State and Catholic Departments. It is the major policy base for the project.

2. Liaison Committee

Each state as well as A.C.T. and Northern Territory has nominated liaison officers who devote time to the development of T.E.P. and who provide a link between the project and evaluation activities throughout Australia.

3. Central Office

Phillip Hughes, Neil Russell and Diana McConachy constitute the professional staff available centrally to implement policy and to initiate appropriate activities nationally.

4. Materials Collection

- . Material relating to all aspects of evaluation from within Australia and overseas is continually being collected. A filing system has been developed and the material indexed. Each article is briefly described on a catalogue card.
- . An ERIC search is being maintained to continue this development further.
- . A series of one-page descriptions of evaluation approaches has been begun and will continue.

5. Activities

- . Liaison established with states through committee and by visits of project staff.
- . Facilitation of particular evaluation studies throughout Australia, on request.
- . Identification of research projects and consultants through bodies such as A.A.R.E.

B. AREAS FOR POLICY AND ACTION

I. RESEARCH

a. Case Studies

If we take the description of a case study outlined by the O.E.C.D.²⁴ which includes both a description of what happens in a given evaluation situation and an analysis of the situation in the light of existing research and theory, then the case study emerges as a significant tool to modify future evaluation practice.

Research which has already commenced includes:

- Analysis of existing descriptive case studies of school-level evaluations to determine the impact (if any) of particular evaluation approaches.
- Commissioning of new evaluation studies to identify the evaluation priorities of different groups (parents, teachers, principals, curriculum advisers).
- Commissioning of case studies where a particular evaluation approach is used (e.g. goal directed) and subsequently analysing the process and consequences of that approach.
- Identification of the important elements of writing a case study and the incorporation of these into a training program.

b. Search for Australian Articles and Research

Whilst a good deal of American evaluation research is available through E.R.I.C., Australian research (including theses and articles) is often not retrievable through a computer network. Considerable effort will be made to seek out Australian research theses in order to produce an annotated bibliography of these materials.

A catalogue of recent evaluation research (1975-1978) reported through the E.R.I.C. system has been completed. However, the review of this research from a variety of perspectives is now required.

c. Descriptions of Evaluation Approaches

One difficulty of the present state of evaluation in Australia is the lack of clarity with which the various approaches to evaluation are described. In many papers (e.g. the article by R. Stake), the description of the process does not readily suggest a practical school based implementation of that approach. There is an immediate need to describe and analyse key evaluation approaches and to develop the suggested implications.

d. Teachers Use of Evaluation Materials

There is a widespread assumption that teachers will have increased responsibility for school evaluation in the immediate future. Research examining the reactions of teachers to curriculum materials in various settings indicates that it is important to explore the nature of evaluation materials that particular schools find most useful. It is proposed that a research project be commenced to investigate this area.

II. DEVELOPMENT

1. MATERIALS

a. Policy Statement

This document provides a rationale for curriculum development in Australia, taking note of developments in the United Kingdom and the United States of America, and comprises three sections:

- Background to the Teachers as Evaluators Project,
- Patterns in the United Kingdom and United States of America,
- Implications for planning and activities in the Teachers as Evaluators Project.

b. Catalogue

Materials available on curriculum evaluation in Australia and elsewhere are constantly being catalogued and filed at the Project Office. Methods for the dissemination of information about these materials are currently under review. An initial step will be to send complete sets of catalogue cards to state liaison officers. Each referenced catalogue card describes the contents of an article.

c. Annotated Bibliography

The compilation of an annotated bibliography is one means of disseminating materials and will be undertaken as an interim measure. The bibliography will be in a loose-leaf form that can be readily expanded to accommodate materials continually being added to the Project's collection.

d. Occasional Papers

It is hoped the production of occasional papers with a theoretical emphasis will draw attention to the broader issues of evaluation, and consequently be of use to a wide-ranging audience.

e. Evaluation Booklets

It is envisaged that the evaluation booklets will be in the form of practical guidelines for school evaluation. They will cater for a variety of needs:

- general introductory evaluation guidelines,
- guidelines for whole school evaluation at primary and secondary levels,
- guidelines for subject/program evaluation.

f. One-page Evaluation Schedules

A series of simplified descriptions of evaluation approaches is being prepared. A diagrammatic representation of a particular approach is accompanied by a one-page prose description which attempts to highlight its advantages/disadvantages. Key references are noted, Australian schools using the approach are cited (school's approval sought), and a suggested time line is set.

g. Evaluation Packages

It is anticipated that these packages will be used in tertiary institutions for general coursework, or for specific pre-service and post-graduate education units. Content of the packages remains to be decided and might include some of the following:

- a one-page evaluation schedule,
- an example of a school using this evaluation approach,
- an evaluation booklet,
- reference to relevant theoretical papers.

h. Workshop Materials

It is intended that workshop materials have two functions:

- to be used in schools to help teachers make decisions about appropriate evaluation approaches,
- to be available for inservice needs.

i. Simulation Film

A well-produced simulation film will have a variety of uses, both in the school workshop situation and in the inservice area. These might include:

- identifying particular groups involved in evaluation,
- creating an awareness of potential difficulties faced by evaluators,
- training facilitators,
- training people to write case studies.

j. Videotaped Series

It has been proposed that people who have been involved in evaluation research/activities be interviewed when possible, and the resultant videotapes will form an evaluation series which will have both theoretical and practical components. Such a series could be used at a variety of levels.

k. Newsletter

It is hoped the newsletter will create an awareness of what is going on in evaluation in Australia. Aimed at a wide audience (teachers, academic staff, personnel from state departments and systems, people displaying an interest in evaluation both in Australia and elsewhere), it will provide a base for the exchange of information. Contributions will come from a variety of sources; including liaison officers and other Project personnel, schools undergoing evaluations, academic institutions, and departmental personnel. This information will present a picture of the evaluation situation as reflected in policy development and ongoing activities.

2. PERSONNEL

Whilst a growing number of people are aware of the need for evaluation at school, regional, state and systems levels, it is becoming increasingly evident that insufficient personnel with evaluation experience are to be found in Australia. Personnel are needed to identify and support ongoing evaluation activities, and the Teachers as Evaluators Project suggests four courses of action to alleviate the situation:

- identifying and preparing those who will train evaluators and facilitators,
- training facilitators,
- training teachers to be evaluators,
- training people to write case studies.

A distinction is drawn between the functions of the 'facilitator' and the 'evaluator'. The facilitator is an external person who is invited by a school to assist in the design and implementation of its evaluation. He does not make judgements about the school. The evaluator may be either from within or outside the school and is involved in the total evaluation process, including the judgemental aspect.

The Summer School to be held in January, 1979 under the guidance of Dr. Wynne Harlen will provide an opportunity for both training of personnel and production of evaluation materials.

3. INFORMATION

Identification of the evaluation policy of each state or system is a fundamental aim of the Project Office.

This information can be used to establish priorities for the co-ordinated production of materials. The Project Office may also be able to provide assistance in states where particular activities might be interrelated to obtain an economy of effort.

At this point it would also be useful to identify current policies, projects, and activities in Australia relevant to curriculum evaluation, and develop some form of model/chart to represent this information. This will help in the co-ordination of various groups contributing to these activities.

C. SCHEDULE OF PROJECT ACTIVITIES

No attempt will be made to schedule Project activities on a time-line until Project personnel meet with state liaison officers in Perth in November. At this meeting strategies will be planned and activities co-ordinated between the Project Office and the states in order to obtain an economy of effort. As an interim measure, Project activities have been ranked in order of a priority and coded according to who will carry them out.

Thus,

- A = high priority
- B = medium priority
- C = low priority

and

- 1 = activities Project Office will undertake itself,
- 2 = activities Project Office can undertake by commissioning/directing co-operating groups,
- 3 = activities which will involve the co-operation of various groups.

I RESEARCH

- A1 Identification of the important elements of writing a case study.
- A2 Analysis of case studies.

- A2,3 Commissioning of new evaluation studies to identify evaluation priorities.
- A2,3 Commission of case studies using a particular approach.
- A2 Teachers' use of evaluation materials.
- B1,2 Search for Australian articles and research.
- B1,2 E.R.I.C. search.
- B2,3 Descriptions of evaluation materials.

II DEVELOPMENT

Materials

- A1 Policy Statement
- A1 Newsletter
- A1 Annotated Bibliography
- A1 Catalogue
- A2,3 Evaluation Booklets
- B1,3 Workshop Materials
- B1,3 One-page Evaluation Schedules
- B1,2 Occasional Papers
- C3 Simulation Film
- C1 Videotaped Series

Personnel

- Al,3 Identifying and preparing those who will train evaluators and facilitators.
- A3 Training facilitators.
- Al,2,3 Training teachers to be evaluators.
- Al,3 Training people to write case studies.

Information

- A1 Identification of state/system evaluation policy.
- Al,3 Establishment of priorities for action.
- A1 Identification of current projects and activities.

ACKNOWLEDGEMENT

In the references acknowledgement has been given to a few of the relevant writings in this area. In addition, Mr. Phillip Hughes wishes to acknowledge the very significant assistance of a number of people through various discussions. In particular, he is grateful for the assistance of Professor W. Taylor, Professor Denis Lawton, and Dr. Helen Simons of the Institute of Education, London University; Dr. Wynne Harlen of the Chelsea College, London University; Dr. Michael Eraut of Sussex University; Dr. Marten Shipman of the Inner London Education Authority; Dr. Robert Stake of C.I.R.C.E., University of Illinois.

REFERENCES

1. C.D.C. Study Group. Curriculum Evaluation, Curriculum Development Centre, Canberra, 1977.
2. Bourke, S.F. and Keeves, J.P. The Mastery of Literacy and Numeracy: Final Report, Australian Studies in School Performance. A.C.E.R., published by A.G.P.S., Canberra, 1978.
3. Husen, T. (Ed). International Study of Achievement in Mathematics, Almqvist and Wiksell, and John Wiley and Sons, Stockholm and New York, 1967. This was the predecessor of a 9 volume series reporting the results of extensive research in six subject areas: Science, Literature, Reading Comprehension, English and French as Foreign Languages, and Civic Education. For a summary see: Walker, David A. The I.E.A. Six Subject Survey. Almqvist and Wiksell, and John Wiley and Sons, Stockholm and New York, 1976.
4. Glass, Gene. The Growth of Evaluation Methodology, Research Paper No. 27. Laboratory of Educational Research, Boulder, Univ. of Colorado, 1969.
5. Hastings, Thomas. Kith and Kin of Education Measures Journal of Educational Measurement 6: 127-130 1969.
6. Tyler, Ralph. Perspectives of Curriculum Evaluation, A.E.R.A. Monograph Series on Curriculum Evaluation, No. 1, Chicago, Rand McNally, 1967.
7. Bloom, A.J. et al. Taxonomy of Educational Objectives. I. Cognitive Domain, David Mackay, New York 1956.
8. Krathwohl et al. Taxonomy of Educational Objectives. II. Affective Domain, David Mackay. New York 1964.
9. Cronbach, L.J. Course Improvement Through Evaluation. Teachers' College Record No. 64, 1963 pp 672-683.
10. Stufflebeam et al. Educational Evaluation and Decision-Making, Phi Delta Kappan National Study Committee on Evaluation, Peacock, Itasea, Illinois, 1971.
11. Scriven, M. The Methodology of Evaluation, in Tyler et al. Perspectives of Curriculum Evaluation, A.E.R.A. Monograph Series on Curriculum Evaluation, No. 1. Rand McNally, Chicago, 1967.

12. Stake, Robert. 'The Countenance of Educational Evaluation'. Teachers' College Record No. 68, 1967, pp 523-540.
13. Assessment of Scientific Development - Report of Assessment of Performance Unit Science Working Group. Department of Education and Science, October, 1977.
14. Accountability in education - the Michigan Experience. Jerome T. Murphy and David K. Cohen, in The Public Interest No. 36 Summer 1974.
15. Wiseman and Pidgeon. Curriculum Evaluation. The National Foundation for Educational Research, 1970.
16. Tawney, David (Ed). Curriculum Evaluation Today: Trends and Implications, Schools Council Research Studies, Macmillan, 1976.
17. Eisner, Elliott. Instructional and Expressive Objectives, in Golby, et al, Curriculum Design, Croom Helm Ltd., London, 1975.
18. Parlett and Hamilton. Evaluation as Illumination, in Tawney (Ed) Op. cit. (14).
19. Stake, Robert E. Program Evaluation, Particularly Responsive Evaluation. Paper presented at a conference on "New Trends in Evaluation", Gotesborg, Sweden, October, 1973.
20. Jenkins, David. Curriculum Evaluation as Literary Criticism. Unpublished Paper.
21. Macdonald and Parlett. Re-thinking Evaluation. Notes from the Cambridge Conference, Cambridge Journal of Education Vol. 3 Easter, 1973.
22. Shipman, Marten. Inside a Curriculum Project, Methuen & Co. Ltd., London, 1974.
23. Worthen, Blaine R. Some Thoughts about Current Models For Evaluating Curricula, Australian Educational Researcher, Vol. 5 No. 3. August, 1978.
24. Case Studies of Educational Innovation: IV Strategies for Innovation in Education. OECD, Paris, 1973.

RELIABILITY OF CRITERION-REFERENCED TESTS

J. F. IZARD.

1.0 INTRODUCTION

The term "reliability" has a variety of meanings. In psychometric usage, reliability is the tendency toward consistency from one set of measures to another, or the internal consistency of a set of measures. In classical test-score theory, observed scores are conceptualized as the sum of two components. The first component, the "true" score, is considered to result from aggregation of systematic factors, whereas the second component, the "error", is considered to be random (Stanley, 1971; Cronbach et al, 1972; Guilford and Fruchter, 1973).

In a given testing situation, there may be intra-individual and inter-individual variability, and in the classical test theory approach, such sources of variation are assigned to "true" or "error" variance according to the purposes of testing. Thorndike (1951) describes a classification of sources of variance for test scores to aid decisions as to whether particular portions of variance are to be regarded as "true" variance or "error" variance.

A number of observations may be made about the concept of reliability as defined in classical test theory, and its