

## CHAPTER 9

# POLITICAL AND ORGANIZATIONAL INFLUENCES ON MIDDLE SCHOOL EVALUATIONS

JAAP SCHEERENS\* and BERT P. M. CREEMERS†

\*University of Twente, The Netherlands

†RION, Institute for Educational Research, University of Groningen, The Netherlands

### Abstract

Middle school evaluations provide a case in point when the influence of political and organizational contextual conditions on evaluations is considered. This is illustrated by means of a description of the experiences with the middle school evaluation in the Netherlands and a brief sketch of the context of similar evaluations in Western Germany, England, Wales, France and Sweden. Several conclusions about favourable and unfavourable contextual conditions are drawn and some conditions concerning the future of middle school evaluations are discussed.

### Introduction

“Political and organizational conditions determine what evaluations can do”, say Cronbach and his associates in a chapter in which they attack the “rational” model of decision-oriented evaluation (Cronbach *et al.*, 1982, p. 76). The experiences with middle school evaluations in Western European countries provide excellent opportunity to analyse this challenging thesis empirically: at this time many European countries have behind them a period of 10 to 20 years of experimenting with comprehensive or “integrated” middle schools. In some of these countries attempts were made to provide conditions for independent evaluation of middle school programmes, according to the logic of quasi-experimentation. As far as theoretical conceptions of the relationship between policy-oriented research and decision making are concerned, we could contrast Cronbach’s instrumental model with Campbell’s notion of “Reforms as Experiments” (Campbell, 1969). In addition to supplying material which bears on important notions in evaluation theory, the analysis of contextual influences on middle school evaluations illustrates some of the workings of educational policy-making when an issue as highly-debated and controversial as middle school development is at stake.

We shall also show that, over time, the emphasis in middle school evaluations tends to shift away from the traditional idea of programme evaluation to other types of evaluation-oriented educational research.

### Frame of Reference

Our frame of reference is a model of “evaluation research in context” developed by Scheerens (1985). According to this model the opportunities for scientific evaluation research are constrained by:

- (1) formal characteristics of the policy-programme that is to be evaluated;
- (2) organizational arrangements, concerning both programme- and evaluation-management;
- (3) the larger societal and political context.

Thus, as an example of the first type of constraint, evaluations are easier to design when programme goals and means have been clearly specified in advance, and when there is general agreement on the overall mission of the programme.

To illustrate the second type of constraint, vital issues of evaluation such as independence and leeway to carry out research activities, are strongly influenced by organizational arrangements, such as the question by whom and according to what procedures the evaluation research is funded.

If we consider the last type of contextual constraint, (larger societal and political context), the degree of (de)centralization can be seen to influence both the nature of educational programmes and the arrangements for evaluating them. At this level, the extent of political debate of the programme is also seen as an important — according to some authors unfavourable — condition for evaluations (Goldstein *et al.*, 1978). Inherent in this way of considering evaluations are three further notions that should be mentioned here. First, contextual influences on evaluations are manifest at all developmental stages of an evaluation, at the stage of planning and design, when building the evaluation team and when conducting the actual information-gathering activities. The final stage, evaluation utilization, is well documented in the literature, though this literature generally neglects contextual influences at earlier stages. Secondly, contextual influences should not be taken as contextual *determinants*: the community of evaluation researchers in a particular country and the surrounding scientific community, must be seen as active agents able to improve conditions for evaluations. In the third place, contextual conditions of evaluations should be depicted as continua and not as black-and-white dichotomies similar to rational policy and experimental evaluation design on the one hand and garbage-can decision-making and qualitative research on the other (for further details see Scheerens, 1985).

This conception of evaluation research-in-context can be used within the framework of evaluation-management and as a tool for meta-evaluations (i.e. meta-evaluations of the review type, concerned more with the way evaluations were carried out than with syntheses of evaluation results). When this latter application is chosen — as we do here — items of description fall in the following categories:

*“Rational structure” of the programme to be evaluated*

(i) the articulatedness of goals and means, uni- versus pluri-finality, hidden policy agendas etc.;

*Organizational conditions*

(i) the way important stake-holders define aims and scope of the evaluation in question and the degree to which they (dis)agree among themselves;

(ii) institutional arrangements concerning the evaluation, such as contractor–evaluator relationships, ways of funding research activities, external vs “in house” evaluation teams;

the status of the evaluators in the scientific community, structures for evaluation management and coordination; relationships between evaluations and practitioner (e.g. teachers);

(iii) organizational issues concerning programme-management; centralized vs decentralized management; a common centrally developed curriculum vs local curricula, etc.

*Characteristics of the educational system at the macro-level*

(i) degree of centralization vs decentralization, extent of autonomy of individual schools, stability of policy vs shifts in governments, characteristics of educational cultures and ideologies at the time of study.

When, metaphorically speaking, these conditions are taken as the “independent” variables in this type of meta-analysis, *quality* — as judged by experts — and *relevance* — as judged by users of results — of actual evaluation-research activities are the attributes on which to measure the “dependent variable”.

## The Case of the Dutch Middle School Evaluation

### *General Orientation*

When in 1973 the Dutch government was formed out of a coalition of socialists and christian democrats, the idea of an integrated first phase of secondary education rapidly gained momentum. A plan for the middle school (grade 7 to 10) was one of the cornerstones of a larger scheme to reform all components of the educational system. The philosophy supporting more equal distribution of educational opportunity lay behind all these innovation plans. Postponement of irreversible choices in educational careers and heterogeneous grouping of pupils were seen as the chief ways of bringing this about. The initial middle school plans added a touch of “reform pedagogy” to these measures towards more equality in education, emphasizing individual growth, social learning and the affective domain of learning. Since there was considerable opposition against the middle school plans from the conservative party, whereas the christian democrats held an intermediate position, a full-fledged implementation of the middle school was unthinkable. So the well-known compromise of an experiment was chosen. From the beginning it was clear that the major political opponents held different views about the meaning of the middle school experiment. The protagonists saw the experiment as the first step in a nation-wide implementation of the middle school. The opponents were inclined to take the idea of an experiment more literally. In 1974 their view was accepted as the official interpretation of the middle school experiment, when the socialist Minister of Education was forced to declare in parliament that the experiment was not just meant to find out *how* to implement a middle school, but that it would also have to address the question of *whether* or not a middle school was at all desirable.

### *Formal Characteristics of the Policy-programme*

As far as formal characteristics of the policy-programme were concerned, the way the middle school experiment was designed and carried out was far removed from the idea of

a scientific experiment. Schools were recruited according to their readiness to innovate. The overall aims of the experiment were never operationalized, let alone specified in terms of measurable educational outcomes. A central middle school curriculum was not developed, the only thing approaching this was a model for a middle school curriculum that was completed years after the start of the experiment and which was still stated in general terms. What actually happened was that each school that took part in the experiment invented its own middle school curriculum.

### *Organizational Conditions Concerning Programme Management*

Structural arrangements concerning programme management enforced the fragmented way in which the middle school experiment was carried out. In fact there was no programme management. Between the central policy planners at the Ministry of Education lay only a semi-formalized network of advisory committees of which the Innovatory Committee for the Middle School (ICM) was the most important one. This committee produced several advisory reports, but had no formal authority to implement their ideas. In fact the ICM had completely adapted its innovatory policy to the reality of autonomous schools and even propagated the idea that each school, as it were, had to invent the wheel "for itself" (critical comments on this innovatory policy were made by Leune, 1981 and Creemers and De Vries, 1982). Although the Dutch educational system is endowed with an elaborate and substantial "support structure" of facilitators, curriculum developers and teacher-educators, these organizations failed to bridge the gap between the central level and the schools.

### *Organizational Conditions Concerning Evaluation Management*

Opinions about the function of evaluation of the middle school experiment differed. The ICM opposed the idea of summative, decision-oriented evaluation. Their preference for formative evaluation was consistent with their view of the middle school experiment as a first step towards nation-wide implementation. Nor was outcome-oriented evaluation high on the priority list of the ICM, which was understandable since they were opposed to the idea of formulating educational objectives (so called "end terms"). In this they were supported by government employees at the Ministry of Education, who said that specifying this type of evaluation criteria was harmful to the innovatory process (Scheerens, 1983, p.99). These government officials said that evaluation should merely provide information on the innovatory process. Schools emphasized evaluations that "they could use", in other words, feedback on what went on in their particular school. The only party involved that publicly defended scientific evaluation-research, including outcome-measurement and the use of a control group, was the Foundation for Educational Research (SVO), which is the central institute that funds and coordinates an important part of educational research in the Netherlands. To coordinate research efforts concerning the middle school experiment, various successive committees designed "research programmes", but an overall evaluation plan was never produced. The actual research activities consisted of a number of scarcely interrelated separate studies, mainly focused on the description of the entrance-situation, the choices of pupils when proceeding

through the middle school, and the description of “where pupils went” after completion of the middle school.

### *Characteristics of the Educational System at the Macro-level*

All organizational arrangements, concerning both programme- and evaluation-management, can only be understood from two interrelated basic characteristics of the Dutch educational system. The denominational freedom of education which has led to what is known as “pillarization” of schools, support-structure and educational pressure groups on the one hand and the autonomy of individual schools on the other hand. This highly differentiated structure has led to a situation where each newly created organ of importance must contain representatives from all pillars. It has been said about this process that it is “good for democracy but bad for getting things done” (cf. De Groot, 1982). The political climate surrounding the middle school experiment changed when in 1977 the socialist–christian democratic coalition was replaced by a christian democratic–conservative coalition. The gradual proliferation of experimental middle schools slowed down. The innovatory committee complained about how slowly the Ministry reacted to its advisory reports. As far as the experimental nature of the middle school innovatory programme was concerned, the ICM as the main advisory committee never took seriously the idea of a “scientific” experiment providing evidence for a “go/no-go” decision. It could be said of the technical motives they put forward for this point of view that they implied a self-fulfilling prophecy: if the ICM would not try to approach the idea of a field-experiment as closely as possible, no-one else would. But apart from these technical reasons the ICM was never reluctant to take a political stance, pleading repeatedly for a nation-wide middle school implementation, when in fact the ongoing experiment had not produced any convincing evidence for such recommendation.

### *The Evaluation Projects*

The actual execution of research projects was handicapped by long delays in obtaining research contracts. Although the SVO (see above) was the immediate funder, the money came from the Ministry of Education, and at that time the Ministry had to approve separately each funding decision. Sometimes SVO and the Ministry disagreed on the funding of a research project, in which case the Ministry usually got its way, after considerable delay. Another difficulty in carrying out the research projects was lack of cooperation from schools. For some schools, the middle school was more like a movement supported by an ideology than simply an organizational and curricular innovation. In particular, intelligence tests and achievement tests were opposed by schools. These technocratic instruments apparently provoked a clash with the humanist-reformist attitude of some of the school teachers. This led to extremely low response rates in several research projects. Because of lack of coordination between individual research projects, it appeared hardly possible to interrelate data bases. Although there were exceptions, most research projects were rated low in quality by experts (Scheerens, 1983; see also Bosker, this issue). The various research projects that were carried out under the banner of middle school evaluation had no apparent effect on the further political decision-making

concerning the revision of the first phase of secondary education. When interviewed (in 1982), key officials at the Ministry of Education said they believed that the research projects in question had hardly any influence on further policy-making. Moreover, later policy plans considering integrated secondary education in the Netherlands, used *ex ante* expert advice (from the Council for Government Policy, W.R.R.) more than the *ex post* evaluation of the middle school experiment.

Summarizing, the following conclusions can be drawn:

(i) although labelled as an experiment, the Dutch middle school programme was far removed from a scientific field experiment: aims were only vaguely specified in advance, schools took part in the experiment on the basis of self-selection and to a large extent had to invent their own innovative curriculum;

(ii) organizational conditions, both concerning programme-management and evaluation-management, were not conducive to an integrated evaluation programme: the programme was strongly diversified, no overall evaluation plan was designed, there was a lack of coordination of the various research projects that were carried out under the flag of middle school evaluation, and individual projects were handicapped because of lack of cooperation from schools;

(iii) factors at the macro-level that lay behind these procedural and organizational conditions were the shift in national policy priorities concerning the middle school, and the position of the Dutch educational support structure within a system that is characterized by the principle of freedom of education;

(iv) the actual research projects were judged low in quality by experts and as low in relevance by decision-makers.

### Political and Organizational Aspects of Middle School Evaluations in other Countries

The Dutch experience with the middle school evaluation strongly supports Cronbach's statement cited at the opening of this article. Before drawing any more conclusions specifically directed at *middle school* evaluations, we shall look briefly at experiences with middle school evaluations in some other West European countries, and consider similar contextual conditions applicable in the Dutch case.

In West Germany the idea of a nation-wide "Gesamtschule" experiment was abolished soon after its initiation in 1969. According to Tillmann (this issue) political differences between the two leading parties made a nation-wide statement impossible. School experiments were carried out and evaluated in the separate states of West Germany. Some of these studies were criticized as not being free from bias, since they were carried out, not by independent researchers, but by state officials.

Other evaluation activities, notably the work by Fend and his colleagues, are considered of high quality. The work of this research group consisted of a nation-wide comparison of comprehensive and existing school types, followed by a similar comparison in three states (Nordrhein-Westfalen, Niedersachsen and Hessen).

Fend emphasizes the independence of his research group (Fend, 1982, p. 18). He is emphatic in stating that his evaluation work did not influence political decision-making on the middle school in Germany. The major decision, namely to have the "Gesamtschule" as a fourth school type in some states, had been taken long before the evaluation results became available.

In France the comprehensive middle school termed the *collège* was enforced by law in 1975. At that time, the conservative government and the opposition agreed on the main aims of comprehensive education. Opposition against the middle school has steadily grown and nowadays its ideas are strongly disputed both in society and in politics (Legrand, 1986, p.52). Legrand (1982) provides a very critical account of the role that educational research played in the political decision-making concerning the structure of the first phase of secondary education. He says with regard to the experiments that preceded the establishment of the *collège* that "when an experiment of this type is begun, the decisions on the changes in question have already been taken by the responsible bodies" (p. 132). From his analysis of the position of the Institut Pédagogique, Legrand concludes that independent evaluation research, where researchers cooperate closely with teachers, is probably incompatible with the highly structured hierarchical system in France. Instead, government officials strongly control evaluation and monitoring activities; furthermore, there are insufficient resources for educational research.

The organizational structure for educational evaluation in Sweden is described by Marklund (1982). Here, each sector of society is governed by a central administrative authority, which operates under the supervision of a particular ministry. The central administrative authority for the educational sector is the National Board of Education (NBE). One of the most important tasks of the NBE is the monitoring and evaluation of educational reforms. The administrative authorities, such as the NBE, are obliged to entrust their policy-oriented research to the universities. The universities thus carry the executive responsibility for policy-oriented research. The whole structure is intended to avoid excessive domination of research by short-term party political interests. From Lundgren's detailed overview of the history of middle school development and evaluation, we conclude that evaluation has played an important role in the development and reshaping of the middle school in Sweden (see Lundgren, this issue). The context of a centralized educational system and a stable government, supporting the educational innovation in question, together with institutional arrangements that call for independent, university-based research, build favourable conditions for evaluations as complex as those concerning middle school experiments.

In Great Britain, the development and evaluation of the comprehensive school differs greatly from the more or less centralized attempts in other countries.\*

In the U.K. the development of the comprehensive school has been a gradual process over the last decades, controlled by the local educational authorities. Here evaluation has taken the shape of monitoring- and assessment-programmes initiated by the central authorities rather than overall programme-evaluation. We feel that these types of monitoring and assessment programmes are a very useful alternative to programme evaluations.

### Conclusion: The Future of Middle School Evaluations

In this paper we have described the contextual conditions of the middle school evaluation in the Netherlands in some detail and included an outline overview of

---

\*Where with respect to West Germany it should be noted that here we are dealing with a national level that is decentralized in the various states, but has strong centralistic tendencies at the state level.

contextual conditions of middle school evaluations in some other European countries. Though the international comparative aspect was perforce superficial, we believe it illustrates the importance of political and organizational influences on evaluations. More specifically, we should like to draw the following conclusions. First, the experiences which we have described show that the beautiful logic of "reforms as experiments" does not seem to be applicable to large-scale national educational reforms. Even when experiments at this level can be realized, as was to some extent the case in Western Germany, the results will come too late to influence major policy decisions. More partial and successive field experiments can still form a powerful evaluation strategy, particularly when the innovation is not under strong political pressure. Secondly, our case study material supports the hypothesis that evaluations stand a better chance, when the political limelight on the programme under evaluation is not too fierce. (We believe this point is illustrated by the experiences in Sweden and Germany on the one hand and the Dutch experience on the other.) This conclusion questions the practicality of certain evaluation methods, where controversy is seen as the very basis of evaluations (e.g. Hofstee, 1985). Thirdly, the material presented shows the importance of independent, university-linked evaluation-institutes as a buffer against political bias (cf. the Swedish experience).

Finally, as by now the more turbulent initial phases of middle school development have been passed in the countries we have described, we believe that both the context and the most likely evaluation approaches are gradually changing. Over the years the sharp edges of integrated secondary education have smoothed out, producing a reconciliation between new and traditional ideas. As comprehensive systems are evolving almost everywhere, all types of internal differentiation are being worked out. The humanist and reform pedagogic ideals have been replaced by a new call for effectiveness and efficiency. Various schools of educational research, including middle school evaluations, emphasize the importance of the school as a level of analysis in evaluation research (e.g. Reynolds, 1987; Gran, 1986). At the same time we see periodic educational assessments and the measurement of educational indicators being developed and applied in many countries. We think that the future of evaluations of the gradual (or partial) comprehensivization in many countries lies in a combination of periodic monitoring and assessment, small-scale and partial experiments and a blending of evaluation research with more "fundamental" educational research directed at school and instructional effectiveness. These approaches may have the added advantage of allowing for accumulated research findings over longer periods, thus relaxing the time pressure of "one shot" evaluations at the end of experimental periods, which, in the opinion of practitioners, are always too short.

## References

- Campbell, D. T. (1969). Reforms as experiments, *American Psychologist*, **24**, 4.
- Creemers, B. P. M., & De Vries, A. M. (1981). Constructie en invoering van de middenschool, *Pedagogische Studiën*, **58**, 357-372.
- Cronbach, L. J. and Associates (1982). *Toward reform of program evaluation*, San Francisco: Jossey Bass Publishers.
- de Groot, A. D. (1982). Research as a learning process. In Kallen *et al.* (Eds.) *Social science research and public policy-making, a reappraisal*, Windsor: NFER-Nelson.
- Fend, H. (1982). *Gesamtschule im Vergleich*, Weinheim und Basel: Beltz.
- Goldstein, M. S., Marcus, A. C., & Perkins Rausch, N. (1978). The non-utilization of evaluation, *Pacific Sociological Review*, **4**, No. 1.
- Gran, B. (1986). The comprehensive schools in Sweden. From primary to secondary education. In W.

- Wielemans (Ed.). *Vernieuwingen in het secundair onderwijs*. Leuven: ACCO.
- Hofstee, W. K. B. (1985). A betting model of evaluation research. In B. P. M. Creemers (Ed.) *Evaluation research in education. Reflections and studies*. The Hague: SVO.
- Legrand, L. (1982). Educational research and policy-making in France. In Kallen *et al.* (Eds.) *Social science research and public policy-making, a reappraisal*, Windsor: NFER-Nelson.
- Legrand, D. (1987). The effectiveness of selective and non-selective schools. Paper presented at the AERA-conference in Washington.
- Leune, J. M. G. (1981). Het innovatieproces Middenschool, *Pedagogische Studiën*, 58, No. 9.
- Marklund, I. (1982). The sectorial principle in Swedish research policy. In Kallen *et al.* (Eds.) *Social science research and public policy-making, a reappraisal*, Windsor: NFER-Nelson.
- Reynolds, D. (1987). The effectiveness of selective and non-selective schools. AERA-paper. Washington.
- Scheerens, J. (1983). *Het sector onderzoek: onderwijsonderzoek in de marge van wetenschap en beleid?* Harlingen: Flevodruk b.v.
- Scheerens, J. (1985). A systems approach to the analysis and management of large-scale evaluations, *Studies in Educational Evaluation*, 11, 83-93.