EVALUATION APPROACHES AND RESULTS IN CURRICULUM RESEARCH AND DEVELOPMENT IN THE NETHERLANDS

Jan van den Akker* and Nico Verloop**

* Faculty of Educational Science and Technology, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands
** Teacher Education Department/Educational Research Center, Leiden University, P.O. Box 9555, 2300 RB Leiden, The Netherlands

Introduction

Curriculum evaluation is a many-sided concept. It links two comprehensive and well-established domains in the field of education, which is illustrated by the fact that each has its own separate volume of The International Encyclopedia of Education (Lewy, 1990; Walberg & Haertel, 1990). The many sides of curriculum evaluation are also reflected by the multitude of curriculum evaluation models presented in the literature (see Alkin, 1994, for a recent overview).

This review article on curriculum evaluation in The Netherlands will focus on evaluation approaches and results in both curriculum research and professional curriculum development practices. Since both constitutive domains, "curriculum" and "evaluation", are somewhat notorious for their elusiveness and the gap between rhetoric and reality, special attention will be paid to the user perspective of curriculum evaluation.

Curriculum evaluation in The Netherlands was reviewed about a decade ago in a special issue of this journal (Verloop & Creemers, 1984), in an article by Creemers and Terlouw (1984). Therefore we will concentrate our analysis on more recent trends and findings. We will refrain from dealing extensively with general, less curriculum-specific forms of educational evaluation, like large-scale (inter)national assessment studies of student achievements (see Hoeben, this issue). However, we will broaden the curriculum scope by including evaluation of less traditional objects and contexts, e.g., evaluation of courseware, and evaluation of training in business and industry.

The review starts with a brief description of the organizational scenery of curriculum evaluation in The Netherlands, distinguishing various aims and contexts of evaluation. Next, we will provide an overview of curriculum evaluation research in the last decade, discussing the methods and outcomes of these studies. In this overview we will distinguish between evaluation research on products from generic curriculum development projects and evaluative findings from other curriculum research. Then, we will pay attention to evaluation practices in different professional development contexts. The article will close with some discussion on recent trends and future challenges for curriculum evaluation.
There are several possible ways of describing the organization of curriculum evaluation:

- A first distinction can be made according to different curriculum contexts, for example: general (elementary and secondary) education, vocational education, higher education, adult education, and corporate training (in business and industry).
- A second aspect concerns the nature of curriculum activities (cf. Walker, 1990): 'generic' (aimed at a large variety of instructional practices in diverse settings, in the Dutch situation usually at a national scale) or 'site-specific' (focused on a specific setting, e.g., a school or training department).
- Third, one may emphasize the type of curriculum at stake (cf. Walberg & Haertel, 1990): products (courses, textbooks, lesson materials; usually focused on planning and delivery of concrete instructional processes), or programs (broader educational proposals and plans for study, somewhat more distant from the direct instructional process, and embedded in more general educational policies).
- Fourth, evaluation activities can be characterized along the well-known distinction between formative purposes (aimed at improvement during development) or summative goals (studying curriculum quality to provide information for decision making, especially at policy levels).
- Finally, curriculum evaluation may be a direct activity itself, or it may be an (important) aspect of some research project in the curriculum domain, in which curriculum evaluation is part of a research design aimed at answering a more fundamental research question.

Since we do not intend to offer a full description of all possible contexts, the description in this section will be limited to the organizational characteristics of the kind of curriculum evaluation research that has been most frequent and is most accessible via publications: evaluation in the context of generic curriculum development for subjects domains in general education. Attention to other organizational contexts will be paid in later sections, when we analyze other curriculum research with evaluative findings, and typical curriculum evaluation practices in various professional development contexts.

Generic curriculum development for general education is a primary task for the National Institute for Curriculum Development (SLO). The SLO (in operation since 1976; staff size: about 150 professional developers) provides services at the request of all kinds of educational bodies and institutes, and of the Ministry of Education and Science. These services consist mainly of the development of model curricula and coordination of curriculum development (cf. van den Brink & van Bruggen, 1990). SLO has carried out more than 200 curriculum development projects for (subjects in) primary, secondary, vocational, and adult education. SLO products (curricular programs and materials) have an exemplary, not an obligatory status. Everyone is free to use these products and to adapt them to their own specific situation and needs. This agreement is based upon political (constitutional freedom of education), economical (relation with educational publishers), and innovation (stimulation of active use) arguments. The organization can be characterized as one aiming at interaction between generic development by a center (SLO) and site-specific activities by schools and teachers. Also, various other organizations, especially educational publishers,
educational advisory centers, and institutes for in-service training, are supposed to play an important intermediary role.

Much evaluation research on SLO products has been initiated via cooperation of three national agencies in the Dutch educational support structure: Institute for Educational Research (SVO), National Institute for Educational Measurement (CITO), and SLO. A special task force with representatives of these institutes was responsible for coordinating evaluation requests and for selecting and funding research groups to execute the evaluation research. Usually the research groups were university-based. Most active in this respect were the Universities of Groningen and Twente; other participating universities were those of Amsterdam, Nijmegen, and Utrecht. During the 80's (especially in the second half) over 40 of such evaluation studies were completed. The findings of these studies will be discussed in the next section.

Review of Curriculum Evaluation Research on SLO Products

The major evaluation question in those studies concerned the impact of SLO projects and their products on educational practice. Several efforts have been made to review, discuss and learn from the entire set or subsets of those evaluation studies (van den Akker, 1993; van den Akker, Boersma & Nies, 1989; van Bruggen, 1987, 1992, 1994a; Marshall, 1987). Below we present a summary of the characteristics and findings of these studies.

a) The type of evaluation questions varied considerably. Various kinds of impact of many types of curricular products (programs and materials) for many different audiences (students, teachers, teacher educators, school counselors, educational publishers, policy makers and administrators) were investigated.

b) The research designs varied strongly in scope and depth. Often a mixture of methods and instruments was used, questionnaires and interviews being the most frequent ones (applied in three-quarters of the studies). In about one quarter of the studies, lessons were observed, and student tests were administered in less than that. In about half the studies, document analysis took place, either as a separate component or by way of preparation for developing other instruments (e.g., observation checklists). The size of the target groups differed widely (from a couple of teachers to 600 students).

c) Not surprisingly in view of this variation, the evaluation results were very diverse. General conclusions were difficult to draw, but it was possible to come up with some very broad statements. Although the majority of audiences was well aware of SLO activities, many products were only moderately known. If respondents did know SLO products, their judgment was in general mildly favorable, especially when they had actually used the products over a longer period of time, and when personal contacts with developers or other participating educators occurred. However, the actual use of SLO products was rather limited in frequency. Moreover, the implementation was often very different from the original intentions of SLO developers. As most users appeared to consider SLO products primarily as a source of ideas, and not so much as guidelines to be followed, this was to be expected.
Drawing more detailed conclusions appeared especially difficult because of ongoing discussions about appropriate criteria for impact of SLO products. Diverse criteria were advocated: reputation of product (or project); degree of product dissemination; impact of product (or series of products from a project) on other educational components (for example: exams, textbooks, inservice courses); user judgments on attractiveness and/or practicality of a product; actual classroom use by teachers, and student achievements.

The appropriateness of the latter two criteria (actual classroom use and especially student effects) for determining the effectiveness of SLO products has been heavily debated. For those criteria it is difficult to determine the differential effect of SLO products, because of the numerous intervening variables that influence the implementation of such products from their original design to their eventual outcomes. Van Bruggen (1987, 1992) has repeatedly argued that the impact of SLO projects and products is not primarily to be found in actual instructional processes and outcomes, but in various more intermediary places: plan of courses as found in schools; central policy regulations with regard to educational contents (e.g., attainment targets, exams); textbooks, teacher guides, and other instructional materials, and professional behavior of teacher educators, test developers, educational advisors, and the like. Since SLO's major task is in developing 'recommended', not 'mandatory' curricula (cf. Glatthorn, 1987), van Bruggen stresses the indirect influence of SLO proposals and materials on taught and learned curriculum.

However, several researchers have taken the position that the ultimate criterion for curriculum impact relates to student achievements (van den Berg, Terwel & Wierstra, 1993; Greners & Hoeben, 1987). Therefore, they tend to concentrate on measuring student results in their evaluation designs, more or less ignoring intermediary effects. Other researchers (cf. van den Akker & Voogt, this issue) have emphasized the central role of the teacher in curriculum implementation. For that reason they tend to focus their evaluation designs on concrete classroom processes so as to analyze the curriculum-in-action.

Besides the discussion on relevant criteria, it should be noted that the discussion about the conclusions has also been hampered by the many methodological problems that often appeared in these evaluation studies. Typical problems related to: valid and reliable operationalization of criteria in instruments (a tendency to over-reliance on self reports); representative sampling of respondents, and dealing with non-response (50% on average).

A last remark should be made about the organization and timing of the evaluation studies. Original evaluation needs and questions of SLO developers were usually of a formative nature: They hoped to receive information that could help them improve their product. However, it often took a lot of bureaucratic procedures and thus a long time before the evaluation questions were translated into requests for proposals, these research proposals were written and judged, funds made available, and studies commissioned. By the time the research had actually been executed, and finally completed with a report, the results usually turned out to be much too late to be of any assistance to the developers. Obviously, the fact that because of these slow processes formative questions usually received summative answers did not stimulate the utilization of the evaluation results (cf. Bonset, 1994).
The last decade has also shown a substantial amount of curriculum research in the domain of general education that was not, or only indirectly, related to SLO projects. A strong impetus for that research came from a kind of special interest group on curriculum research, sponsored by the Institute of Educational Research (SVO). Participating researchers from different universities discussed their research plans and results within a commonly designed framework that strongly reflected the implementation thinking that was so influential in the 80s (cf. Fullan, 1982; see Snyder, Bolin & Zumwalt, 1993, for a historical analysis of different perspectives). A major impetus for this international trend had come from a review by Fullan and Pomfret (1977) of research on innovation efforts in curriculum and instruction, revealing rather massive failure of previous RD&D projects with their typical adoption perspective. The emphasis on implementation implied more attention to the context, needs, perceptions, and (re)actions of innovation ‘users’, especially teachers. Also, it elucidated the different dimensions of curriculum innovations for individual teachers: not only using other materials, but also changing teaching approaches and educational beliefs. Above all, thinking in terms of implementation rather than adoption meant viewing educational change as a process with many variables at issue, instead of as an event with immediate outcomes.

However, this common research interest in implementation did not cause strong convergence in evaluation methodology between Dutch curriculum researchers, as will become clear from the brief overview of research projects below.

At the University of Groningen a number of evaluation studies were executed comparing the effects of different teaching courses or textbooks on student outcomes. The relevance of student achievement as a central criterion in curriculum evaluation was emphasized by Creemers and Hoeben (1987) and Hoeben (1987) [see Hoeben, this issue]. Different subjects were investigated: van den Berg (1987) did his research in the domain of economics, Harskamp (1988) focused on arithmetic, van Batenburg (1988) on mother tongue, and Edelenbos (1990) on English.

In general, these studies could hardly detect different effects of different curricula on student achievement. However, some comments on these findings seem appropriate. First, substantial effects of curriculum products (especially when they are innovative) cannot be expected unless their implementation is supported by other forms of assistance (see, e.g., Fullan, 1991, for an extensive review of such strategies). Second, it often appears that different products, although their design was inspired by different curricular conceptions and ideals, lose much of their innovative aspiration during their development by commercial textbook publishers. Moreover, many textbooks lack sufficient procedural specifications for teachers to support them with concrete lesson preparation and execution. The usual tendency of teachers to adapt innovative proposals to more familiar approaches, decreases even further the chance that differences in actual instructional processes and resulting student outcomes will occur. Finally it must be noted with regard to this type of comparative evaluation that often the tests used do not adequately reflect the specific innovative characteristics of
the different curricula under study, which increases the chance of masking potential differential effects (a problem already discussed by Walker and Schaffarzick, 1974).

The Groningen researchers themselves too are of the opinion that curricular products do matter for instructional planning and classroom learning. They emphasize the importance of systematically incorporating principles of effective instruction into the materials (Creemers, 1991; Hoeben, this issue).

In the research approach of several researchers at the University of Utrecht much attention was paid to student reactions and achievements. Terwel (1989; see also Terwel & van den Eeden, this issue) did some consecutive studies on the effectiveness of maths curricula (secondary education) that emphasized cooperative learning. Wierstra (1990; see also Wierstra & Wubbels, this issue) did evaluative research in the context of a long curriculum development project for physics (secondary education) that emphasized context- and activity-based learning.

Over the years, both researchers were more closely linked to the development work in their domains than in the case of the Groningen researchers who merely acted as external, summative evaluators. Their strong interest in 'progressive' forms of learning in their projects (and probably also their methodological preferences; cf. van den Berg, Terwel & Wierstra, 1993) focused their evaluation measurements on student outcomes. In both research lines only limited effects on the student level could be traced. Only few and indirect data on the teacher level were gathered.

Curriculum researchers at the University of Twente tended to pay specific attention to the implementation perspective of teachers. Being involved in various design-oriented studies (or forms of 'development research', see van den Akker & Plomp, 1993), these researchers built on the curriculum typology of Goodlad, Klein and Tye (1979) to distinguish between curriculum representations of original designers ('ideal' and 'formal' curriculum), teachers ('perceived' and 'operational') and students ('experiential' and 'attained'). They saw it as essential to focus curriculum evaluation on what teachers think and actually do in their classroom (perceived and operational levels) in order to understand the frequent gap between designers' visions and student outcomes. An important evaluation procedure for measuring actual lesson processes was the use of innovation and practice profiles (see van den Akker & Voogt, this issue).

Also typical for most of these studies were the iterative revision and improvement of draft versions of curriculum materials during formative evaluation before more summative evaluation took place. This final testing occurred either via an experimental approach (van den Akker, 1988, on teaching materials for elementary science; and Keursten, 1994, on teaching materials in geography courseware) or a case study design (van der Geest, 1991, on courseware for writing instruction; and Voogt, 1993, on courseware for physics).

The main aim of this series of research projects was to generate a set of validated design principles for curriculum (including courseware) designers that deliberately take into account the implementation perspectives of teachers (van den Akker, 1994; van den Akker, Keursten & Plomp, 1992).

Besides these primarily product-oriented evaluations, several researchers paid attention to other factors that seem conditional for effective curriculum implementation. For example, Snippe (1991), Vankan (1991), as well as Roes and van den Akker (1993) paid attention to the contribution of in-service education to curriculum implementation. Their focus was somewhat different. Snippe focused on forms of classroom consultation
(for teachers of primary maths), Vankan on training teachers to develop their own materials (for environmental education in secondary schools), and Roes and van den Akker on the use of curriculum materials as demonstration and practice tools within in-service scenarios (for various subjects in secondary education).

Also helpful for understanding the (lack of) impact of curriculum development projects are the investigations of Stokking and Leenders (1990) on the process of dissemination of the huge amount of information that schools receive from the various support agencies in our country. They revealed, for example, that only one-fourth of the target group reads the information sent to them, and that only one of three of those readers actually does something with that information in his/her teaching practice. Such findings underline the complexity of efforts to stimulate educational change on a wide scale.

Another example of curriculum evaluation in a broader innovation context can be found in the work of Slavenburg and colleagues (Slavenburg, 1986; Slavenburg & Peters, 1989). They investigated the effectiveness, in terms of student achievement, of courses for different subjects in elementary schools as part of a comprehensive intervention program on compensatory education, including family involvement. Their rigorously designed evaluation studies underscored the difficulty of realizing significant progress in student achievement in such improvement projects.

A last example of curriculum research with substantial evaluative strength is the comprehensive study of Kuiper and Alting (1990) on the teaching practices in different science subjects in lower secondary education. Their extensive survey (with questionnaires completed by 856 teachers) offered the kind of representative base-line information on school and classroom practices that is so rarely available at the start of curriculum reform efforts. In addition, Kuiper (1993) provided an in-depth description and analysis via case studies (including 75 lesson observations) of exemplary teachers, focusing on the context- and activity-based quality of their classroom practice. The studies urged for a greater sense of reality, in the light of the gap between the actual and the desired situation. Various implications for more effective curriculum design and implementation in this domain could be derived.

Some separate remarks should be made on efforts to evaluate curricula at the meso level of the school or institution (in Dutch: schoolwerkplan). Van der Werf (1988, 1993) showed that effects of such programs on student achievement are hard to trace. However, she claims some evidence that the use of a well elaborated schoolwerkplan can further the effectiveness of the school organization and classroom instruction. At the same time, a study by Lentz and van Thyll (1989) revealed much confusion about the central function of the schoolwerkplan. Recent approaches seem to emphasize the integration of schoolwerkplan development and evaluation into systematic efforts of quality assurance in schools (Hoeben, 1993; Janssens, 1993).

Evaluation Approaches in Development Practices

Another perspective on curriculum evaluation is to leave the work of researchers behind, and to focus on the evaluation practices of developers in different professional contexts. The main concern in this section is to analyze how professional developers actually practice (formative) evaluation approaches in their daily tasks. We start with a
description of the role of evaluation in different contexts before reaching some general conclusions.

For the context of generic curriculum development, quite some information is available from a study by van den Akker, Boersma and Nies (1990; see also van den Akker & Boersma, 1993) on development strategies of SLO projects in the past fifteen years. That evaluative study (mainly consisting of a series of 18 reconstructive case studies) was initiated because of the need for reliable and systematic knowledge of actually practiced development strategies in order to learn from previous experiences and to improve professional methods, procedures, and techniques for curriculum development. The need for this improvement was underscored by the results of several evaluation studies on the impact of SLO projects, which raised serious doubts about the effectiveness of the development activities (cf. van Bruggen, 1987).

The study offered an overview of typical patterns in SLO projects. With regard to the evaluation component in the development practices the following pattern emerged:

- Formative evaluation activities were usually carried out by SLO developers themselves. Draft products were mostly evaluated in a single round. Products on the micro level (materials) were more often evaluated, and with a greater variety in methods, than products at more abstract levels (programs).
- The emphasis in the evaluation approach was on acceptability (attractiveness, consent) and utility (practicality, feasibility). Effectiveness (in terms of impact on student reactions and achievements) was rarely addressed.
- The primary evaluation focus was on content matter; pedagogical aspects received less attention.
- Teachers (usually from pilot schools) and SLO colleagues served as most important sources for evaluation.
- The most frequently used data collection methods were: discussions at project conferences with teachers and many other different participants in the curriculum innovation process (e.g., teacher educators and school advisors), and written comments on draft product versions by various stakeholders. More direct and structured data collection in the school and classroom practice (by interviews, observations, and teacher logs) was seldom used.

A general conclusion was that the role of formative evaluation in SLO projects should be improved. Although this conclusion has been underlined in several small scale follow-up studies on SLO practices (e.g., Bonset, 1994) some persistent obstacles (especially a lack of time and expertise) still seem to hinder a satisfactory realization of that intention. However, this is not a particular problem of SLO but rather a common characteristic for comparable curriculum development agencies in other countries (van Bruggen, 1994b).

When we look at evaluation practices in other contexts, somewhat similar patterns emerge. Although no extensive research-based information is available, different sources (Pieters & Bergman, 1994; Voerman, 1993) suggest that in the development context of corporate training the actual attention for (formative) evaluation is less than suggested by models in the literature (e.g., Brinkerhoff, Brethouwer, Hluchyj, & Nowakowski's 1983 guidelines, are well-known among Dutch training developers) and also fall short of what the developers themselves express as desirable (cf. Kessels, Smit & Keursten, 1992). Although lack of time and resources is the most frequently mentioned reason for this, it also seems that developers'
commitment to evaluation is often insufficient for actual evaluation activities. The belief that evaluation should not only investigate the impact of training programs, but also help improve impact by being an integral part of the training development process is still not common (Voerman, van den Akker, & Keursten, 1994). In this domain too, the findings seem to correspond with practices in other countries (see, e.g., Wedman & Tessmer, 1993, for a comparable study in the USA).

The interest of corporate management in evaluation is increasing, too, but their concern is usually focused on cost-effectiveness (cf. Mulder, 1992). For that purpose, van Sandick and Schaap-Neuteboom (1993) developed an instrument which they applied in a quasi-experimental study to establish the returns on investment of job training for the middle management of a large retail business.

Comparable findings seem to hold for the domain of courseware development. Although formative evaluation usually does occur, especially in nationally sponsored development projects (Collis, Kanselaar & Moonen, 1990) and appropriate instruments are available (e.g. Kanselaar & de Tombe, 1991), constraints of time and financial resources appear to hinder systematic and intensive evaluation procedures (Moonen, 1989). As a result many courseware products lack sufficient clarity and practicality (Keursten & Nieveen, 1992). Guidelines for a formative evaluation approach that is integrated in the design and development process of courseware, have been formulated by Keursten (1994), who paid special attention to the characteristics of teacher materials for inexperienced computer users.

The actual developments and evaluation strategies in the context of educational textbook publishing are rather difficult to reconstruct. Hardly any research data are available (cf. Venezky's, 1992, conclusion for the US situation). The only relatively recent Dutch study on the practices of educational publishers (de Groot, 1988) offers descriptive information at a rather general level. The amply available informal data suggest that systematic formative evaluation, going beyond expert and consumer appraisal and including field tests, is rather limited, probably because educational publishers don't feel the commercial need for such in-depth investments (cf. Elliott & Woodward, 1990, for comparable US findings). Publishers' emphasis in evaluative activities tends to be on the orientation stage of preliminary study of context and needs analysis.

However, there certainly is public interest in the quality of textbooks. The role of textbooks in stimulating (or hindering) a recent large-scale curricular reform across all subjects in lower secondary education is especially manifest. Also, there is a recent initiative of the Netherlands Information Centre for Educational Resources [NICL] for more thorough summative procedures for analyzing and assessing the quality of textbooks. Results from such studies should support teachers in evaluating textbooks for potential selection.

A general conclusion about the position of evaluation in professional design and development practices must be that evaluation is one of the first activities to be neglected if resources (especially time) are scarce. There seems to be constant tension between intentions and reality: less attention to evaluation is often being paid than proclaimed by the developers themselves, and certainly less than prescribed in the extensive literature on design and development models. The perceived relevance of evaluation is apparently modest, and efficient integration of evaluation activities in professional routines has so far been unsuccessful.
Discussion and Trends

A general conclusion from the (mainly summative) curriculum evaluation research previously described (both related to SLO products and not) was that the effects (on whatever level) were usually not very substantial. Those rather consistent and disappointing outcomes, combined with the fact that they often were published too late to have any formative consequences, have led to a decision to stop the common SLO/SVO/CITO funding of such summative studies. Nevertheless, the studies contributed to greater awareness and understanding of the many contextual and intervening variables that potentially influence the process and outcomes of curriculum dissemination and implementation (cf. van den Akker, 1993; van den Berg, Hameyer, & Stokking, 1989; van Bruggen, 1992; Fullan, 1991). This has underlined how tricky it is to try to assess the effects of a single curriculum product on a large scale: Methodological and practical problems are numerous, and the rather uninformative results seem hardly worth the larger investment.

Currently, we observe two trends that promise to be more productive. The first is the strong increase in different kinds of large scale evaluation and assessment studies which do not try to determine the specific effect of a single product, but aim at more comprehensive analysis, evaluation and monitoring of the state of the art in a subject in school practice. Three complementary kinds of studies can be mentioned:

- Periodic assessments of subjects, started in the second half of the 80s in elementary education, and organized by CITO (see also Hoeben, this issue). Almost all subjects have been evaluated this way; follow-up studies, providing opportunities for longitudinal comparisons, have begun.
- Active Dutch participation in studies of the International Association for the Evaluation of Educational Achievement [IEA]. Recently completed studies are, for example, those on ‘Reading Literacy' and ‘Computers in Education'; ongoing is the ‘Third International Mathematics and Science Study' [TIMSS]. Although the IEA studies focus on student achievement ('attained' curriculum), they also pay attention to data that represent the 'intended' and 'implemented' curriculum, thereby increasing the chances for relevant curricular analysis.
- And, since a few years, national evaluation studies by the Inspectorate, analyzing school and classroom practices in various subjects, based on a large number of school visits, including document analysis, interviews and lesson observations.

Taken together, these studies can help to build a more complete and longitudinal picture of instructional processes and outcomes in the educational practice at large, including the impact of curriculum innovation projects over the years. Also, they can be of help to curriculum policy makers and developers by pointing to problematic curriculum areas that need improvement. Another advantage of these studies is that they create the opportunity for linking evaluative data from the micro level of instruction with data at school and national levels, thereby providing directions for policy measures that stimulate the quality of education (Hoeben, 1993; Janssens, 1993).

An important challenge for the future is to optimize the interaction between these different types of studies in order to improve the interpretation and utilization of their results.
The second trend relates to the limited utilization of (formative) evaluation as discussed in the previous section on development practices. Several recent initiatives reflect the desire for more integration of formative evaluation in patterns of curriculum (including courseware and training) development. This trend fits within a general tendency towards more rational development strategies, illustrated by components like: specification of product aims and target groups; standardization of routine tasks; utilization of scientific knowledge; more empirical validation of product quality; anticipation of implementation problems, and attention to cost-effectiveness. More systematic inclusion of formative evaluation in development processes should not only result in better products and implementation assistance, but also in more opportunities for 'strategic' learning for different participants in the process, expanding and sharpening their knowledge base for future development tasks (cf. van den Akker & Plomp, 1993).

Sufficient evaluation expertise in development teams and the availability of procedures and instruments that combine informativeness and efficiency, are necessary for adequate evaluation activities. Obviously, the development context is important in this respect. For example, SLO recently produced evaluation guidelines, including exemplary instruments, for the curriculum developers. Currently, SLO is also involved, in a joint project with the University of Twente, in an exploratory study on the design of a computerized 'performance support system' for formative curriculum evaluation (Nieveen, van den Akker & Plomp, 1994). An interesting characteristic of this system (called CASCADE) is that the nature, focus, and methods of evaluation evolve during the development process. 'Screening' or 'Initial Debugging', and 'Expert Appraisal' of draft products are, for example, most prominent in early development stages, while 'Micro Evaluation' and 'Try-outs' are typical for later stages, and 'Field Tests' are most appropriate for a final stage. In such an (iterative) approach the distinction between formative and summative evaluation loses some of its sharp edges, especially when field tests with parts of the larger target group are built into the project (cf. the integrated evaluation approaches for various professional domains of van den Akker & Keursten, 1993; Keursten, 1994; Tessmer, 1993; Thiagarajan, 1991).

Some interesting comparable efforts for computer assistance in evaluation activities are going on in other contexts. In the domain of corporate training, for example, Rosendaal and Schrijvers (1994) designed a performance support system for course designers (called COCOS), including support for evaluation tasks. Moreover, Jansen and Weide (1993) have developed a computerized support system for evaluation in higher education, focusing on designing and analyzing the results of student questionnaires.

No doubt these improvement efforts are increasingly driven by the push to quality assurance and cost-effectiveness that permeates not only the world of corporate training, but almost any professional field. That drive may create incentives for researchers to contribute to the design and validation of such evaluation methodology and technology. If these research activities and findings lead towards less rhetoric and more quality improvement in development practices, then substantial progress will be made.
References


Evaluation Approaches


Keursten, P. (1994). *Courseware-ontwikkeling met het oog op implementatie: de docent centraal* [Courseware development from an implementation perspective: A central role for the teacher]. Enschede: University of Twente.


The Authors

JAN VAN DEN AKKER is Associate Professor in the Department of Curriculum, Faculty of Educational Science and Technology, University of Twente, Enschede, The Netherlands.

NICO VERLOOP is Professor of Education at Leiden University, The Netherlands, and Director of the Leiden Educational Research Centre (BOVO).