

THE EVOLUTION OF COMPUTER-RELATED EDUCATIONAL EVALUATION IN THE NETHERLANDS

B. Collis,* G. Kanselaar† and J. Moonen*

**Department of Education, University of Twente, Enschede, The Netherlands*

†Faculty of Social Sciences, University of Utrecht, The Netherlands

Computer-Related Educational Activity in the Netherlands

Computer-related educational activity in The Netherlands can be traced from two perspectives - from that of its development through the independent, exploratory activities of individuals or groups, or from the perspective of government-supported activity. The latter context can be seen as emerging as early as 1978, when a government advisory committee was established to formulate recommendations with regard to the social consequences of information technology in education (Rathenau Advisory Committee Report, 1980). The comments of this report relative to education "aroused the interest of Parliament in information technology and education" (van Deursen, 1987). This interest, accompanied by a recognition of the value of a more clearly defined central government policy in this area, has led to various large-scale nationally funded projects in The Netherlands.

For example, in 1981 the Dutch Minister of Education announced the creation of a national centre for education and information technology (called the COI), to be located at the University of Twente in Enschede. The main tasks for this centre were to inform, to coordinate, and to stimulate the use of computers in schools. Because using courseware is a major activity in this context, one of the main responsibilities given to the COI was to inform schools about available courseware and its value for education.

Another initiative began in 1983 when the Dutch parliament discussed a policy paper submitted by the government concerning the stimulation of information technology in The Netherlands. A motion was approved in which Parliament asked for evaluation criteria for educational software in order to raise the quality of software used in Dutch schools. As a result, the Dutch Institute for Educational Research supported a research project for the development of an instrument to evaluate educational software. The research was executed at the University of Utrecht and resulted in a report and a software evaluation instrument which became available in 1985 (Kanselaar et. al., 1986).

In 1984, the Dutch government launched, as a major policy initiative, the five-year Informatics Stimulation Plan (Informatica-Stimuleringsplan, 1984). Within this plan many activities were planned, including the setting up of a regular service concerning educational software description and evaluation. A project called INKON was defined, to be executed by two institutions, one of which was the COI. The COI was asked to evaluate software packages available for Dutch education. This led to the development of a separate project within the COI, called SCEN (Software and Courseware Evaluation Centre for The Netherlands). Drawing upon the work done by the Utrecht project, SCEN developed its own evaluation instrument and started in 1986 publishing a regular journal, circulated to all Dutch schools, about the results of its evaluation activity.

Another activity initiated under the INSP program involves a series of collaborative projects at three so-called "experimental schools," secondary schools chosen to be exploration sites related to computer use. Among the different projects operating in the context of the experimental schools is COMPRO. COMPRO, which is still in progress, is examining "effective and efficient implementation" strategies for school use of computers, focusing on how the teacher can be better supported in his or her attempts at implementation (van der Klauw, 1988). Part of COMPRO also involves software evaluation, this time using a different evaluation instrument and procedure than those in earlier projects.

In the projects which have been described, evaluation has focused on a procedure to assess an already-existing product. A different approach to the evaluation of software is also occurring in the context of the POCO Project. POCO began in 1987 as a national software development plan (POCO Plan of Action, 1987). An evaluator was assigned to this project from

its start, so that continuous feedback can be supplied to the project, not only about its internal functioning but more importantly as to how to improve the likely reception which the software products will have when they are eventually evaluated by individual teachers in terms of "use" or "not-use" decisions in their own classrooms (Collis & Moonen, 1988).

As a final example of a national investment in evaluation relative to computers in education, the educational programme of the INSP Project has also been recently evaluated (Scheerens, 1988). Although this evaluation was only commissioned near the end of the scheduled five years of the INSP Project, it has the potential of influencing a follow-up project in The Netherlands, called PRINT, which began in 1989 and which will oversee the subsequent support of projects related to information technology in schools.

These various national projects that involve evaluation either as a focus of or a tool for computer-related activities in education can be considered more closely in terms of the insights which have been gained from them. In particular, we see two major trends evolving through these experiences. We would like to discuss five of these projects - the Utrecht Courseware Evaluation Project, SCEN, COMPRO, POCO, and the INSP evaluation - in more detail relative to these trends.

Trend 1: Towards a Broader Context for Finished-Product Evaluation

A definite evolution can be seen in the evaluation perspectives employed in the Utrecht Project, SCEN, COMPRO and POCO. This is the realization that product evaluation needs to be considered within the context in which the product is most likely to be used, rather than being an analytical exercise that can be done through looking at software as an isolated product. This is a perspective which has not yet become standard in many settings outside The Netherlands so a close look at its evolution here, in the context of the four projects, is valuable.

The Utrecht Courseware Evaluation Project

In 1984 a courseware evaluation instrument was developed. The evaluation part of the instrument consisted of three components:

- the perceived integration of the courseware package in the curriculum and in the classroom;
- the didactic design of the program
- the technical design of the program

The total instrument consisted of fourteen clusters of items and a total of 112 items. The reliability of the instrument was established by asking thirty judges to evaluate twenty-three courseware packages. It was difficult to find judges who had some experience of courseware and to find courseware packages which were worthwhile to evaluate in 1984. The reliability of the instrument depended very much on the type of software. The judges differed also with regard to the didactic qualities of simple drill-and-practice programs. Extensive explanations and illustrations were written for the judges but also for teachers who were interested in the use of courseware in the classroom. The instrument was aimed at product evaluation of courseware packages by a professional evaluator working with a computer at his desk. In 1984 it was too early to evaluate the implementation of courseware packages in the classroom at a national level.

SCEN

There developed a growing awareness that not only the "intrinsic" characteristics of a software product determine its value, but that the other aspects add to the perceived value of the product when it is used in practice. The SCEN procedure for software evaluation extended the focus of software evaluation by building in end-user comments from teachers who had worked with a software product in actual classroom settings. SCEN began the approach of locating teachers to serve as prospective evaluators, giving the teachers some training in the use of a common evaluation framework, and then using the reactions of those teachers after they had used a product in practice as part of the evaluative input. This approach yields at least some information about the software in the broader context of the classroom - information which can only be predicted in approaches such as the earlier courseware evaluation project. There are definite limitations to the approach, however; for example, the perspective given is

heavily influenced by the personal characteristics of the teacher and class who experienced the use of the software. Also the process begins at the point where a teacher already has a software package, rather than considering the larger context in which the teacher first becomes aware of a software selection, makes a choice of which package to try, and is able to obtain the package for use.

COMPRO

The COMPRO approach to software evaluation includes this larger perspective. The approach corresponds to a gradually developing awareness in the field in general that computer use is not a unidimensional phenomena, but is always influenced by many different contextual variables, all of which also interact with each other (Clark, 1985; Collis, 1988; Ely & Plomp, 1986). The COMPRO approach therefore focuses more specifically on questions the teacher should consider in anticipation of the implementation realities that he or she is likely to confront during the use of a particular programme in the classroom. The COMPRO project in general, although not specifically in the context of its software evaluation component, also considers the relationship between the overall school environment and organizational patterns and the type of computer use that occurs. A new project, just getting underway in The Netherlands, will consider the sorts of organizational support, at both the school level and beyond, that may be particularly effective in helping teachers make good evaluation decisions about software selection and use. Both projects also consider the usability of the evaluation procedure itself by the teacher. The balance between conciseness (and thus likelihood of being used by busy practitioners and validity is another aspect of the evaluation procedure that needs consideration.

POCO

The POCO project extended the approach to software evaluation beyond the context of its implementation setting. Within POCO, the evaluator looked for predictors of end-user reaction, both teacher and student, on an ongoing basis throughout the development process. A major strategy in this prediction process is to arrange "confrontations" between teachers and software at a number of points in the design and development process (POCO Plan of Action, 1987). Therefore, a perspective for the evaluator of the POCO project was to assess how well this "confrontation"

process was functioning. More importantly, the evaluator made suggestions on an ongoing basis as to how this feedback loop could be better realized; thus, the contribution of the POCO evaluation was formative, rather than summative.

From these examples, we see in The Netherlands a clear pattern of evolution with regard to strategies and perspectives for the evaluation of educational software. The pattern could be visualized by a set of concentric circles, in which each embedding circle includes the perspectives of circles within it, but also attempts to bring more attention to the larger context in which educational computer software will function. The evolution in evaluative perspectives parallels, as was noted earlier, the growing awareness in the international research community that computer impact in education is a multifaceted activity. Computer use in the "real" world of the classroom does not happen out of context (Collis, Walker & Grant, 1987); The Netherlands' experience suggests that software evaluation strategies should and can reflect this contextual complexity.

Trend 2: Toward Ongoing Evaluation of Computer-Relative Activity in Education

The POCO evaluation described in the previous section is also of interest in that it illustrates a second aspect of emerging perspectives regarding evaluation of computer-related educational activity - the aspect of systematic, ongoing formative evaluation of complex processes as opposed to evaluation as a summative activity. The INSP evaluation was not able to operate in this fashion, relative to the INSP in itself, as it was only given its evaluative mandate within the last year of the lifespan of the project. However, the evaluation team recommended that its work be seen as pro-active, rather than retroactive, "with an overall formative orientation, i.e., to yield information that could help in shaping the ongoing decisions in informatics stimulation" (Scheerens, p. 10), utilizing an approach that was "improvement-oriented rather than accountability-oriented" (p. 17). The conclusions of the study reflect this orientation.

The POCO evaluation, because it was so commissioned, involved systematic input from the evaluator from the start of the project. Eight reports, totalling approximately 300 pages, were submitted to the director of the COI, as the person with formal accountability for the project but without

specific involvement in its management, over the first eighteen months of the project. The design of the evaluation was adapted from the "contingency-congruence" model of Robert Stake (Stake, 1973; Collis & Moonen, 1988). The approach involves both responsive flexibility, important in a long-term project involving many participants, and "preordinate" focusing, in this case, the continuing appraisal of how likely it was that POCO would meet the overall goals originally established for it, considering the ongoing modifications of the project relative to execution (Schermerhorn & Williams, 1979). The approach also involves continuous feedback loops relative to the "logical contingencies" of the project as it was planned (rather than as it was executed). An assessment of lack of congruence between project expectations and project execution can be constructively analyzed using this approach, in that mismatches (or lack of congruences) which occur can be diagnosed either as symptoms of "logical" miscalculations in the project expectations or as examples of implementation difficulties which are amenable to more immediate reaction, or a combination of both (Shapiro, 1985). Such a framework facilitates two different sorts of ongoing, pro-active evaluative feedback: suggestions that relate to fine-tuning or for alternative strategies for parts of the project which are not questioned in themselves; and suggestions that relate to more fundamental reconsideration of project assumptions and planning. The two aspects of feedback both appeared to make a positive contribution to the POCO project as it evolved during its first eighteen months.

Generalizing from the Dutch Experience

The Dutch experience with respect to evaluation of national-scale, computer-related educational activities has identified two major trends with respect to such evaluation: respecting the broader context in which computer use will happen, and incorporating systematic, ongoing formative evaluation into both product development projects and project frameworks, such as the INSP. We believe that these trends can be productively generalized to the work of evaluators of computer-related educational activity outside The Netherlands. We have already noted, relative to the first Dutch trend, the consistency between the "broader context" orientation for evaluation and the corresponding broader perspective on overall research regarding computers in education which is emerging in practice. We can also identify some particular reasons to support the generalizability of the second trend: the inclusion of systematic, ongoing formative evaluation in computer-related projects and other activities. These reasons are:

1. Computer-related educational activities are difficult to evaluate because of the rapidly changing nature of the medium itself and of what educators are learning about the potentiality of the medium in the classroom.
2. Computer-related educational activities are often difficult to evaluate because goals and objectives of such activities may themselves change during the lifespan of an activity (for the same reasons as in Point 1), or because the goals may be stated in vague and hopeful terms ("Students will be better prepared for the future", "Education will be revolutionalized") that are not possible to use as evaluation criteria.

We believe that systematic, ongoing evaluation can help in the process of goal clarification and adaptation (Baker, 1988), and at the very least can serve as a trace of the motivations for any change of goals or objectives that occur. We believe this contribution can be substantial.

References

- Baker, A.L. (1988). *Sensitive technology assessment of ACOT*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.
- Clark, R.E. (1985). Confounding in educational computing research. *Journal of Educational Computing Research*, 1 (2), 137-148.
- Collis, B.A. (1988, July). *Manipulating critical variables: A framework for improving the impact of computers in the school environment*. Paper presented to EURIT '88, Lausanne, Switzerland.
- Collis, B.A. & Moonen, J. (1988, April). *Designing an external evaluation of a large-scale software evaluation project*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.
- Collis, B.A., Walker, D. & Grant, W. (1987, June). *Not to be taken out of context: Research on computers in the classroom*. Paper presented at the National Educational Computing Conference, Philadelphia.
- Ely, D.P. & Plomp, T. (1986). The promises of educational technology: A reassessment. *International Review of Education*, 32 (3), 231-250.
- Informatica-Stimuleringsplan, Onderwijsbijlage. (1984). Den Haag: Staatsuitgeverij.

- Kanselaar, G., Vossen, P., van de Perel, R., Havekes, F. & Stevens, F. (1986). *Courseware nader bekeken: Eindverslag van het SVO-project 1086: "Voorwaarden voor coursewaregebruik"*. Den Haag: SVO, Selectareeks.
- POCO Plan of Action. (1987). Enschede: COI.
- Rathenau Advisory Committee Report. (1980). *Maatschappelijke gevolgen van de Micro-electronica*. Den Haag: Staatsuitgeverij.
- Scheerens, J. (1988). *The evaluation of the educational section of the Dutch Informatics Stimulation Programme (INSP)*. Paper in press, *World Yearbook of Education* (1990).
- Schermerhorn, G.R. & Williams, R.G. (1979). An empirical comparison of responsive and preordinate approaches to program evaluation. *Educational Evaluation and Policy Analysis*, 1 (3), 55-60.
- Shapiro, J.Z. (1985). Evaluation for a work site programme in health, science and medicine: An application of Stake's model of contingency and congruence. *Educational Evaluation and Policy Analysis*, 7 (1), 47-56.
- Stake, R.E. (1973). The countenance of educational evaluation. In B. R. Worthen & J.R. Sanders (Eds.), *Educational evaluation: Theory and practice* (pp. 106-125). Belmont, California: Wadsworth Publishing Company.
- van der Klauw. (1988). *Proefstations in het voortgezet onderwijs*. The Hague: Informatica Stimuleringsplan, Cluster V.
- van Deursen, K. (1987). The introduction of information technology in the Dutch educational system. In T. Plomp, K. van Deursen and J. Moonen, (Eds.), *CAL: for Europe: Computer-Assisted Learning for Europe* (pp. 3-7). Amsterdam: North-Holland.

The Authors

BETTY COLLIS is an Associate Professor in the Division of Educational Instrumentation Technology at the University of Twente. She is active in various evaluation projects involving technology in education and training and has a particular interest in methodologies for assessing the impact of computer-related innovations in different cultural and instructional settings.

JEF MOONEN is a Professor in the Division of Educational Instrumentation Technology at the University of Twente. He is active in policy and planning involving technology in schools, with a particular focus on educational software development. He is also involved in cost-effectiveness research and methodology relative to computer-related innovations in education and training.

GELLOF KANSELAAR is a Professor in Instructional Psychology in the Division of Educational Research at the University of Utrecht. He is active in various research and development projects involving technology in education and training. He has a particular interest in the construction of Intelligent Tutoring Systems.