Evaluating Electronic Performance Support Systems: A Methodology Focused on Future Use-In-Practice

Betty Collis and Carla Verwijs, *Faculty of Educational Science and Technology, University of Twente, The Netherlands*

SUMMARY

Electronic performance support systems, as an emerging type of software environment, present many new challenges in relation to effective evaluation. In this paper, a global approach to a 'usage-orientated' evaluation methodology for software product is presented, followed by a specific example of the application of the methodology to electronic performance support systems. The usage-orientated evaluation methodology is characterised by: (a) its applicability throughout the life-cycle of a software product; (b) its blending of theoretical and practical aspects through a consistent focus on the eventual outcome, 'Will they use it?' ('use in practice'); (c) its use of a core evaluation instrument, the 'usage map', in various ways during the life-cycle of the product; and (d) its support of decision making for both those involved in the design process of the product and those representing its potential users.

CONSIDERATIONS FOR AN EVALUATION METHODOLOGY

An electronic performance support system (EPSS) presents many interesting problems from an evaluator's perspective. This section briefly discusses some of the issues that have to be considered.

Evaluation of EPSS: general considerations

Need for an evaluation methodology. There is no need here to justify the overall importance of evaluation in the life-cycle of any electronic product; the literature on this is extensive and can be found, expressed in different ways, in many different disciplines including software engineering (see, for example, Ratcliff, 1987), human–computer interaction (Hix and Hartson, 1993) and multimedia educational software development (Reeves and Harmon, 1993). EPSS, as complex and costly systems whose characteristics and boundaries as a software category are as yet only vaguely understood (Collis and Verwijs, 1995), clearly will benefit from systematic, communicable evaluation insights, which in turn require an appropriate evaluation methodology.

Problems in the evaluation of EPSS. Even if the need for evaluation of EPSS is obvious, the methodology for it is not. One problem is that many evaluation approaches focus on how well the use of the system (compared to alternative systems) helps the user to reach a specific performance goal. With EPSS, however, the great variety of ways and contexts in which a user will turn to the system for support make it very difficult to isolate the system as a causal factor in that user's performance or to compare one system with 'an alternative'. The interlinkage of components of an EPSS, including its chosen functionalities, architecture, user-interface aspects, content, and organization and indexing of content, also make evaluation of the system as an entity elusive. And the critical aspect of the organizational factors that constrain or facilitate the user's access to the EPSS '. . . *on demand, at the moment of need* . . . ' (Gery, 1991, p. 34) must also be integrated into any evaluation approach; adding complexity and further making it difficult to select an evaluation methodology from existing possibilities.

General problems confronting any evaluation methodology. In addition to the problems that accompany the evaluation of EPSS (due to their inherent characteristics), there are also problems that confront any evaluation of software products. These aspects relate to the evaluation's execution and usefulness in

practice (Collis, 1989). Two major categories of these aspects are:

- evaluation is frequently not done, or not done early enough to influence the design of a product; and
- those in decision making roles concerning the product must listen to the evaluation and consider its recommendations.

In the following section we describe a general evaluation methodology reflecting the above considerations that we have developed for evaluating complex electronic performance support systems, including not only EPSS but also other sorts of environments such as those supporting collaborative work and computer conferencing. After the description of our general methodology, we briefly illustrate how we are applying it to an EPSS project.

A USAGE-ORIENTATED EVALUATION METHODOLOGY FOR ELECTRONIC SUPPORT ENVIRONMENTS

Roots of the methodology

By 'methodology' we mean an integrated system of theory and procedure for conceptualizing and carrying out an evaluation and interpreting its results. Perspectives which have strongly influenced our evaluation methodology include:

- The need for evaluation to occur throughout the life-cycle of a product, the evaluation having different functions at different points in this life-cycle (Scriven, 1967; Stake, 1973; Nielsen, 1993; Hix and Hartson, 1993; Tripp and Bichelmeyer, 1990).
- The need for frequent interactions with representatives of the end-users of a product throughout its life-cycle, particularly through the technique of rapid prototyping (Nielsen, 1993; Hix and Hartson, 1993; Tripp and Bichelmeyer, 1990).
- The value of an orientation throughout the life-cycle of a product toward likelihood of eventual user acceptance of the product (Dickerson and Hedman, 1993).
- The need for an easy-to-understand and easy-to-carry-out evaluation instrument and procedure if representatives of end-users are to be frequently involved in critical discussions about the product, particularly in the early stages of the life-cycle of the product (Hix and Hartson, 1993, MacLean, Bellotti and Shum, 1993; Rettig, 1994).

We classify the life-cycle mentioned in the above four points as consisting of four phases as described below.

Phase 1. Iterative conceptualization of the product and agreement among the design team and representatives of potential users as to what the product should do, be like, and how it will be used (this phase can be called that of 'design–space analysis', see MacLean, Berlotti and Shum (1993), although design–space consensus is perhaps a better term for our approach).

Phase 2. Iterative clarification of the design through rapid prototyping (Hix and Hartson, 1993; Rettig, 1994; Tripp and Bichelmeyer, 1990).

Phase 3. Beta versions of the product, in a form ready for limited field testing and formative evaluation of the product (Scriven, 1967; Nielsen, 1993).

Phase 4. The public version of the product, along with its accompanying documentation, support, and method of presentation to the potential user. Here a baseline index for evaluation is purchase and a second is use.

The first three of these phases will overlap and often backlap, as, for example, early 'lo-fi' (low-fidelity; Rettig, 1994) rapid prototyping can sharpen consensus about the design space. Similarly, later versions of rapid prototypes may also be beta versions of the product, but may also cause the process to loop backward for further revision, both in terms of new prototypes and also in design–space clarification. Design–space consensus will continue to build throughout the process. Usability testing and user interface testing begin early in the life-cycle process and continue throughout, in different manifestations.

Overview of the 'Usage-Orientated Evaluation Methodology'

Given these roots, and reflecting the considerations noted in the introductory section of this paper as critical for any evaluation methodology, our 'Usage-Orientated Evaluation Methodology' can be expressed as shown in Figure 1. The methodology can be seen as having a 'framework' aspect, relating to the orientation and contextual positioning of the evaluation; an *'activities'* aspect, relating to how the evaluation is executed, when, and with what instruments and procedures; and a *'purpose'* aspect, showing how the output of the evaluation activities is put to use.

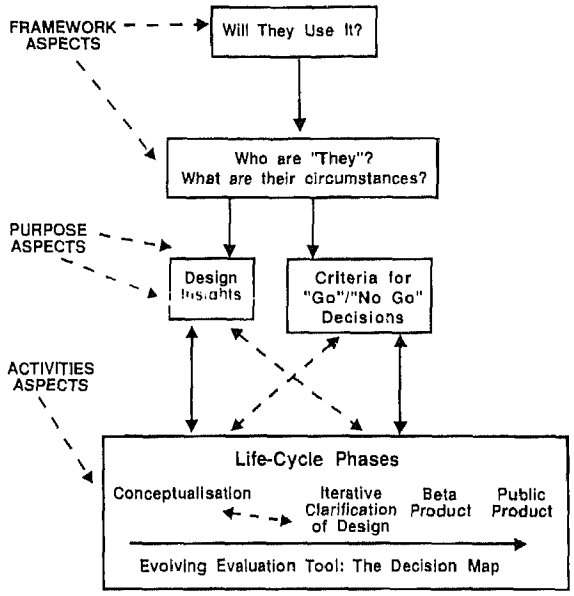

Figure 1. *Overview of the Usage-Oriented Evaluation Methodology*

The Framework Aspect

The components of the Framework Aspect are its orientating principle and theoretical basis, and its contextual filter.

Orientating Principle. The orientating principle for our methodology is usage. We have argued elsewhere (Collis and De Vries, 1993) that 'predicting use in practice' is a strong choice for an evaluation focus for a number of reasons:

- such prediction can occur at any point in the life-cycle of a product;
- if a product is not eventually used, then it will not be commercially viable, and no critical mass of support and insight about the product will develop;
- it has the benefit of being an orientation that is easily understood, yielding a measure which is intuitively compelling to decision makers.

Predicting use in practice will not yield a measure that is possible to obtain with precision; it will stay a guess, but it should be a well-informed guess. Thus knowing the context is also critical to the framework aspect of our evaluation methodology.

Contextual Filter. After usage is accepted as a key orientation for the evaluation, identifying the context of the usage must be the next dominant aspect of the methodology. Techniques from human factors analysis (Hix and Hartson, 1993) with regard to user modelling and representation, and Stake's technique of elaborating and regularly reconsidering existing assumptions about potential users and their circumstances (Stake, 1973; Collis, 1993) are of influence here.

Key Ideas. In simple terms, the key ideas of the Framework Aspect are:

- keep the evaluation focus throughout on 'Will they use it?'; and

– continually refine your understanding of the context.

The Purpose Aspect

Two components. In order for the evaluation to be meaningful, it must deliver valuable insights in a way and form that is seen as useful by those in decision making positions about the product. In our methodology, we represent this by the two components that comprise the 'Purpose Aspect' of the methodology. One of these components has as its purpose to contribute to decision making about the design of product at critical moments in its life-cycle. The other of the components has as its purpose to contribute evaluation criteria relevant to the product for decision makers. We call the first the *'design insights'* and the second the *'go?/no-go? criteria'*. The design insights reflect the spirit of Scriven's formative evaluation (1967) and the technique of rapid prototyping (Tripp and Bichelmeyer, 1990); the 'go?/no-go' criteria are more 'summative' in their nature.

Key ideas. In simple terms, the key ideas of the Purpose Aspect are to use the evaluation experiences at any point in the life-cycle of a product to answer the following questions:

– how can we improve the design so that eventual usage is more likely to occur?
– how can we decide if a product will be used in our setting?

The Activities Aspect

The Activities Aspect relates to the four phases in the life-cycle of a product and their accompanying evaluation strategies, but expresses the evaluation in terms of a standard approach, terminology, and analysis tool.

Relationship of evaluation techniques to life-cycle phases and to the Purpose Aspects of the methodology. Against the four phases of the life-cycle of a product evaluation activities have different emphases but continue to inform both components of the Purpose Aspect. During the design–space analysis, evaluation is primarily focused on helping the participants to structure and interpret their exploration of the design space. However, the evaluation should also steer a preliminary analysis of the factors most likely to affect the eventual use in practice of the product.

During the iterative design-clarification phase of the product life-cycle, the critical factors likely to influence eventual usage are further used to provide structure and interpretation to the reactions of user representatives to the various versions of the rapid prototype of the system.

During the beta testing and public phases of the product, extensive design modifications are less or not possible, but a usage orientation can still benefit the product development life-cycle by contributing ideas about the support of the product in practice in order to anticipate contextual hurdles to usage which the product is likely to encounter.

The 'Decision Map'. Important to all this evaluation activity is having an instrument that is easy to use, facilitates communication between the evaluator and those making use of his insights, facilitates interpretation of evaluation data, and which requires relatively little cost and time in its application. Thus we feel that having a single form of instrument to use in an ongoing way throughout the evaluation is particularly helpful to facilitate these communication and interpretation aspects. We also feel that the instrument should focus all who use it toward the orientating principle of the evaluation, to improve its interpretability.

After many different experiences with many different types of data-gathering instruments, we are currently investigating a type of one-page, paper-based instrument that we think is particularly helpful in making our usage-orientated evaluation methodology manageable. The idea was stimulated by the work of Dickerson and Hedman (1993) and involves showing the various factors felt to be critical in influencing the eventual usage of a product in practice as nodes along paths radiating in an inverse-tree fashion up to the final node of 'usage'. Nodes earlier in a path are predicted as having particular influence on those following in the path. The decision map thus serves to organize the factors that need most attention relative to their effect on eventual usage and to suggest a sequence in attention to those factors.

Key Ideas. The key ideas of the Activities Aspect are thus:

– decision making about eventual use in practice is rooted in a user profile;
– the passage from the 'User Profile' to the prediction of use in practice can be thought of as radiating upward along various paths, each of which has nodes which sequentially influence each other; and

– a negative value in one node suggests a negative ripple effect up through the decision map unless another node can exert a sufficiently strong positive force to overcome the negative effects.

We believe the decision map instrument is relatively simple to use, and communicable to decision makers (thus relating to the critical considerations we noted earlier with regard to evaluation methodologies). Simply stated, if a certain variable is a powerful influence on usage of a product, and if that variable is negative, then why bother spending time and attention on factors that are secondary to it?

Evolution of the decision map during the different phases in the life-cycle of the product. The decision map for a particular context will not spring forward in its final form at the beginning of the evaluation/life-cycle activities. Its evolution, in consultation with others involved in the design and development process and with decision makers representing the end-users, is an important product of the evaluation methodology itself. The decision map should be easy to modify, as new insights emerge concerning the design space of the product. It can focus discussion and facilitate communication and through this process will and should be repeatedly revised. Thus its validity is a direct function of its own development process.

APPLYING THE USAGE-ORIENTATED METHODOLOGY TO AN EPSS CASE

We believe that the usage-orientated methodology is particularly appropriate for EPSS. The orientation toward use in practice is intrinsically relevant to EPSS; the outcomes to be associated with the use of an EPSS are often not well-specified at the beginning of its design (Witt and Wager, 1994); the involvement of many different actors in the EPSS life-cycle makes a common, consistent, communicable evaluation instrument particularly important; a lack of consensus and clarity is highly likely to characterize EPSS design in its early phases; and the interrelated purposes of evaluation relative to informing the design process and supporting the commitment-to-use process are well suited to products whose value is rooted in their being used to support context-specific tasks.

To illustrate this we briefly discuss an EPSS project in which we are involved where the usage-orientated methodology is being applied. In this project, we are partners in the design and development of an EPSS to support instructional designers in a large training organization in their choice and combination of media and technologies for courses and lessons that they develop. Although partners in the process, we are also expected to steer and conduct the evaluation of the EPSS. The details of this project are not given here but can be found in various internal reports (Verwijs, 1994). We will call the situation the Design Case.

The Design Case and the Usage-Orientated Evaluation Methodology

The Framework Aspect of the Evaluation Methodology: relating the usage orientation to the evaluation. The contract for this project stated that the EPSS to be developed would be considered successful to the extent to which it would eventually come to be used to support the practice of the instructional designers in the training organization. Thus the orientating principle of our evaluation methodology related directly to the use-in-practice goal of the decision makers commissioning the project.

The Framework Aspect of the Methodology: analysing the context. We have spent considerable time getting to know the working procedures of the training organization commissioning the EPSS as well as the more specific procedures of its instructional designers. A major task of our context analysis has related to finding the persons best able to illuminate for us the ways of working of the instructional designers in our context and the types of support they would most likely wish to obtain from an EPSS.

Our most important observation was that the instructional designers in this particular centre were highly trained in a certain procedure for specifying lesson planning (a procedure which is supported for them by the 'Handbook for Instructors' which they use as a common format for lesson specification). This procedure asks the designer to indicate clearly and in a distinctively structured way five categories of information and planning: describing the students, the objectives, the location of the lesson, the media to be used, and the positioning of the course in the yearly calendar.

Given this as a starting point, we came to the conclusion that the EPSS to be designed in this project would most likely be accepted and understood, at least initially, in practice if it served as an environment specifically based on support of the designers in completing this required (and appreciated) lesson-plan document. From a three-phase categorization of EPSS (Collis and Verwijs, 1995) that we had

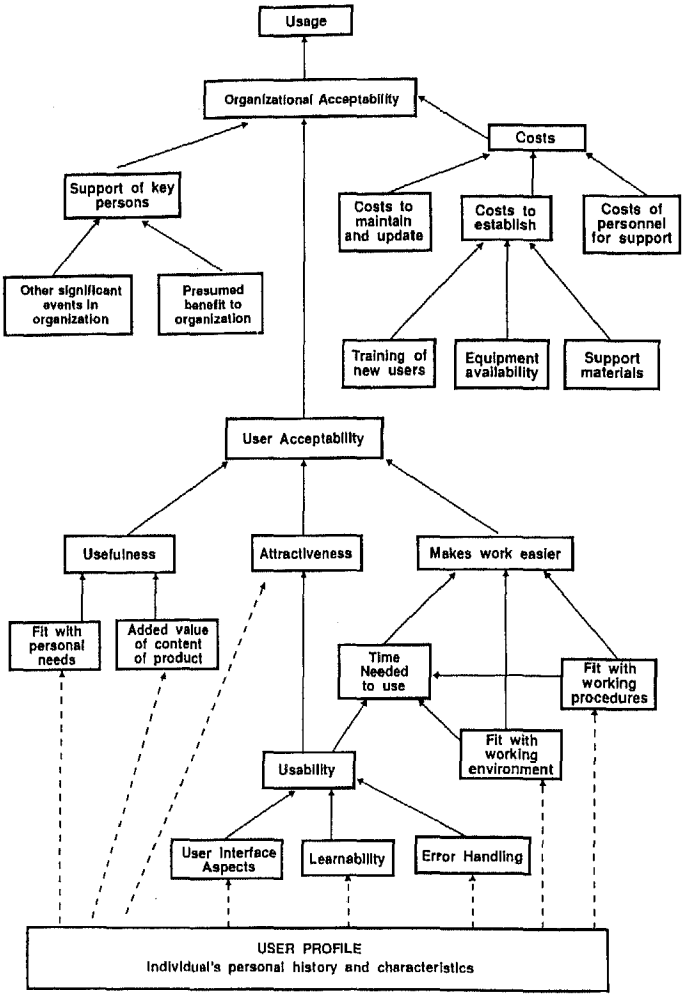

Figure 2. *Decision map for an EPSS*

identified earlier, we provisionally decided to focus the EPSS predominantly on the 'focusing and coming to a particular solution phase' category of EPSS and secondly, and much less dominantly, on the 'browsing for new ideas' category of EPSS. Also, we formed a first version of the decision map for the evaluation, indicating which variables appeared from our contextual analysis to be most likely to influence subsequent use in practice of the EPSS. Figure 2 shows this first iteration of the decision map for the Design Case EPSS.

In the process of determining these initial variables and drafting the first iteration of the decision map, we moved from the Framework Aspect of the evaluation methodology and into the Activity Aspect.

The Activity Aspect: supporting design–space analysis. Typical of many EPSS projects (Witt and Wager, 1994), a major challenge in our project is the initial lack of a shared consensus of what the EPSS should in fact be and do. 'Supporting instructional designers in their technology choices and

decisions' was easily agreed upon, but how this might be done and in what form the support should be offered is more difficult to develop as a shared vision. To better clarify the design space itself we decided to move into a first round of rapid prototyping.

The first rapid prototype. The first rapid prototype is clearly a 'lo-fi' paper prototype (Rettig, 1994) indicating a horizontal view of the EPSS's functionalities and a first suggestion of the user interface's screen design. The latter is built around a visual representation of the lesson-plan document with which the instructional designers are familiar but expressed only in an abbreviated, schematic form. Associated with different steps in the lesson-plan are options that can lead to more and various kinds of information about media and technology choices associated with a particular part of the lesson-plan. This information resides 'behind' the lesson-plan metaphor, and is a hyper-organized multimedia data base of information and examples of different sorts of technologies being used in different sorts of instructional settings. These settings are indexed to correspond to choices in the lesson-plan. This can provide structured support for the 'browsing aspect' of the user's task – getting ideas about different possibilities of media and technology choice relevant to situations similar to his own. Similarly, we argued the value of other functions in our prototype, particularly tools to support cost-effectiveness predictions (see Collis and Verwijs (1994), for a lengthier description).

This first prototype served a valuable purpose, as it quickly suggested clarifications and changes not only to us as designers but to our colleagues from the training centre. As a mechanism of communication, the paper-and-pencil version was revised as soon as it was put forward – which was its intention. Thus the current ideas of the EPSS are evolving rapidly, as iterative refinements of this initial sketch.

The Purpose Aspect: putting the activities to an evaluation purpose. As we continue the process of getting reactions to these iterative lo-fi prototypes from various persons representing the intended users of the EPSS and of those making decisions about its further development, we are using our decision map instrument (Figure 2) to structure our discussions and our own thinking about the discussions.

Thus there are multiple results emerging from this activity loop:

- a step forward toward a shared view of the design space for the project and what is meant by an EPSS for this situation;
- a next iteration of the decision-map instrument to capture our view of what variables need most attention relative to eventual usage;
- ideas for a next round of rapid prototyping; and
- feedback to the Purpose Aspects of the evaluation, primarily as a contribution to decision making about the design of the product relative to eventual use in practice, but also a contribution toward relevant criteria for decision making about the product itself.

REFLECTIONS ON THE EVALUATION METHODOLOGY

Thus we express our Usage-Orientated Evaluation Methodology as being:

- based on the prediction of 'Use in Practice' and thus addressing the question, 'Will they use it?';
- rooted in knowing the potential user and his or her context;
- involving iterative loops of feedback to iteratively emerging rapid prototypes beginning at the level of the design space itself; and
- visualized throughout by a 'decision map' which helps identify barriers to eventual usage.

So far, this usage-orientated methodology seems to provide a useful tool for the evaluation of complex support environments such as EPSS. (We discuss some other applications in another document; see Collis and Verwijs, 1994). We believe that the guiding orientation of predicting use in practice is a very appropriate one for EPSS and for communicating about the evaluation with others. We also think the decision-map instrument has good potential as a socially constructed view of the usage domain. However, many iterations of the predictive process with a range of persons well-acquainted with the context in which the use-in-practice has to occur, with the decision map as an organizer. This is in itself another strength of the methodology if it stimulates interaction and a striving towards consensus among a variety of representatives of the prospective users as well as the design team.

REFERENCES

Collis, B. (1989) Assessment of educational programs: How can we assess 'Dynamics and Vision'? Paper prepared for the International Customer Exec-

utive Seminar, IBM International Education Centre, La Hulpe, Belgium (March).

Collis, B (1993) Evaluating instructional applications of telecommunications in distance education. *Educational and Training Technology International*, 30(3), 266–274.

Collis, B. and De Vries, P. (1993) The emerging Trans-European Network for Education and Training: Guidelines for Decision Makers. Report produced for the Commission of the European Community under Contract 92-001-NIT-109/NL. Task Force Human Resources, Education, Training and Youth, Brussels.

Collis, B. and Verwijs, C. (1994) A usage-orientated evaluation methodology for EPSS. Internal report. Faculty of Educational Science and Technology, University of Twente, Enschede, The Netherlands.

Collis, B. and Verwijs, C. (1995) A human approach to electronic performance support systems: Hybrid EPSS. *Educational Technology*, 35(1), 5–21.

Dickerson, K. and Hedman, L. (1993) Usability and standards. In Byerley, P.F., Barnard, P.J. and May, J. (eds) *Computer, communication and usability: Design issues, research, and methods for integrated systems* (pp. 413–452). Elsevier, Amsterdam.

Gery, G.J. (1991) *Electronic performance support systems*. Weingarten Publications, Boston.

Hix, D. and Hartson, H.R. (1993) *Developing user interfaces: Ensuring usability through product and process*. John Wiley and Sons, New York.

MacLean, A., Bellotti, V. and Shum, S. (1993) Developing the design space with design space analysis. In Byerley, P.F., Barnard, P.J. and May, J. (eds.) *Computers, communication and usability: Design issues, research, and methods for integrated systems* (pp. 197–219). Elsevier, Amsterdam.

Nielsen, J. (1993) *Usability engineering*. Academic Press, Inc., Boston.

Ratcliff, B. (1987) *Software engineering: Principles and methods*. Blackwell Scientific Publications, Oxford.

Reeves, T. and Harmon, S. (1993) Systematic evaluation procedures for instructional hypermedia/multimedia. Paper presented at the Annual General Meeting of the American Educational Research Association, Atlanta, GA (April).

Rettig, M. (1994) Prototyping for tiny figures. *Communications of the ACM*, 37(4), 22–27.

Scriven, M. (1967) *The methodology of evaluation*. AERA Monograph Series on Curriculum Evaluation, No. 1. (pp. 39–89). Rand McNally, Chicago.

Stake, R.E. (1973) The countenance of educational evaluation. In Worthen, B.R. and Sanders, J.R. (eds.) *Educational evaluation: Theory and practice* (pp. 106–125). Wadsworth Publishing Company, Belmont, CA.

Tripp, S.D. and Bichelmeyer, B. (1990) Rapid prototyping: An alternative instructional design strategy. *Educational and Training Technology International*, 38(1), 31–44.

Verwijs, C. (1994) Year One of the Project 'An Environment for the Support of the Selection and Combination of Media for Education and Training': Interim Report. Faculty of Educational Science and Technology, University of Twente, Enschede, The Netherlands. (In Dutch)

Witt, C.L. and Wager, W. (1994) A comparison of instructional systems design and electronic performance support systems design. *Educational Technology*, 34(6), 20–24.

BIOGRAPHICAL NOTES

Dr Betty Collis is a faculty member at the University of Twente in The Netherlands. She received her PhD in measurement and evaluation of computer applications in education in the early 1980s and since then has been continuously involved in evaluation projects involving innovative applications of computer-related technology in education and training. Drs. Carla Verwijs is a PhD student working with Betty Collis on the design of an EPSS for the support of instructional designers in their selection of media.

Address for correspondence: Betty Collis can be contacted at the Faculty of Educational Science and Technology, University of Twente, Postbus 217, 7500 AE, Enschede, The Netherlands. Email: COLLIS @EDTE.UTWENTE.NL