

Evaluating Trans-national Tele-learning Demonstrator Projects: Design and Methodology

Betty Collis and Jolande Vingerhoets, *University of Twente, The Netherlands*

SUMMARY

What is an effective methodology for the evaluation of trans-national projects whose purpose is to demonstrate some aspects of telecommunications-supported distance education, or tele-learning? By their nature, such tele-learning demonstrator projects frequently involve technologies new to many of the participants as well as many organizational complexities given the number of partners – course providers, educational institutions in different countries, telecommunications providers, and educational specialists – involved. This paper suggests a methodological framework and design features for the evaluation of such projects built upon the assumption that finding a key focus for the evaluation, of interest to all of the major stakeholders, is a critical step. In addition, the value of a responsive rather than a preordinate model or the evaluation as well as the choice of case-study methodology is argued. A generic evaluation plan is developed, involving on-going multiple case studies transected at regular intervals by cross-case analysis events. The plan is illustrated by its current application to a major trans-national tele-learning project, TeleScopia, involving courses developed by six course providers, adapted for trans-European offer via a variety of applications of advanced telecommunications.

INTRODUCTION

In this paper we address the question: what is an effective methodology for the evaluation of broad-scale trans-national tele-learning demonstrator projects? We begin with an overview of major aspects of such demonstrator projects, with example references to projects in the DELTA Programme of the Commission of the European Community and a particular project, TeleScopia, taking place from 1994–95. After a brief summary of some predictable difficulties facing evaluation projects relating to educational innovations in general, we look more specifically at some additional problems confounding complex phenomena such as trans-national tele-learning demonstrator projects, with their multiple partners and multiple contexts. From this we move to a suggestion for a methodological framework and design features for such projects expressed in terms of: (a) choice of a purpose and focus for the

evaluation, based on the target audiences for the evaluation; (b) choice of a general model for the evaluation; (c) choice of a general strategy or method for the evaluation along with a consideration of some of the technical issues accompanying this choice; and (d) design decisions within the chosen method for the evaluation including planning for data capture and analysis. Finally, we give an example of the application of such a framework and evaluation plan in the context of the TeleScopia Project.

TRANS-NATIONAL TELE-LEARNING DEMONSTRATOR PROJECTS

In Europe, as in many other regions of the world, large-scale trans-national tele-learning demonstrator projects are being conducted. In order to define what we mean by this type of project, each of the key phrases in the above sentence needs elaboration.

Large scale is easy to understand but harder to define within a preset definition; in general by large scale we mean projects involving many different players: different institutions, organizations, technical providers, funding agencies, and even different researchers and evaluators, each with their own set of local agendas. *Trans-national* is similarly easy to understand in its basic meaning but increases in complexity and variation relative to the difference in languages and differences in culture among the players. *Tele-learning* means that in one or more ways telecommunications is involved in the learning process. This can occur in an increasing variety of ways: adding interaction and communication possibilities to one-way broadcasting or other distance-teaching delivery strategies; allowing real-time collaboration or tutoring between participants at a distance from one other; allowing asynchronous discussion, debate and communication; conveying a tele-lecture from one location to distributed and responsive learners; allowing access to distributed learning resources and other forms of information and expertise; and combinations of any and all of these (for a good review, see Van den Brande, 1993). *Demonstrator projects* are those which include as a major goal the demonstration of the possibilities of a new entity, in the case here, of trans-national tele-learning. Demonstrations are aimed for particular audiences: the entity is being demonstrated to some groups (usually including those who fund the demonstrator project) for some purpose, often relating to further decision making. The defining aspect of a demonstrator project is that it is a specific intervention operated in a real-life field setting, an innovation operated at or near full-scale in a realistic environment (see Glennan, 1987). However, demonstrator projects also imply that the activity involved is not part of on-going practice, but still at the project or intervention level, and thus involves extra infusions of personnel, support and other variables (with tele-learning projects, these typically include new telecommunications infrastructure access, new hardware and software, new resources and sites for contacts, and a timeline and budget that run over and across the normal work of the participants in their non-project situations).

Demonstrator Projects as Subjects of Evaluation

Demonstrator projects have a long history in education and have been the object of evaluations through much of that history (see for example,

Madaus *et al.*, 1989). The purposes of demonstrator projects have been globally categorized as policy implementation, policy formulation, and knowledge development (Yin, 1993, 57). Implementation demonstrators take existing ideas already supported by policy and attempt to put them into everyday practice. Formulation demonstrators attempt to show that new ideas can be feasible in a field setting, enough so that more controlled implementation and supportive policy should be planned. Whatever type of demonstration, the experiences are expected primarily to produce knowledge and insight about the eventual implementation of an innovation in non-project settings. This in turn means that the implicit or explicit question of generalizability is addressed in demonstrator projects: Could an ordinary agency or institution carry out the programme with fidelity without special project support? Under what conditions? (Yin, 1993, 57; Berk and Rossi, 1990, 52).

Tele-Learning Demonstrator Projects and Their Evaluation

Tele-learning demonstrator projects have many variations. Many of the 22 component projects of the Commission of the European Community's DELTA Programme (Telematics Systems for Flexible and Distance Learning) for the period 1991–94 could be called tele-learning demonstrator projects, but with many different manifestations and focuses including the design, production and use of learning materials within tele-learning scenarios; issues relating to pedagogy and communication support as well as training and monitoring support; and scenarios for tele-learning in themselves, particularly relative to their cross-cultural marketability (see 'DELTA at Work', by the Tribune Consortium, 1994, for a comprehensive summary). The technologies involved in DELTA provide a good summary of those available for other tele-learning demonstrator projects in the early 1990s including:

Relating to communication

1,200–14,400 modems, datapack, X.25 and X.400 protocols, ISDN and PSDN services, 2 MB highspeed lines, telefax, telephone, satellite broadcasts (I:BS and video), mainframe networks, LANs, e-mail services, computer conferencing systems, video conferencing, desktop multimedia conferencing, ordinary post, and additional distribution of learning materials via CD-ROMs, laser disk and paper.

Relating to media (digitized and non-digitized)

ASCII text, SGML/GML, graphics, pictures/images, presentation sequences, animation, sound, motion pictures, video still pictures, interactive hypermedia, computer software, computer applications such as word processors and other tools, paper-based media, analogue audio, and analogue motion (VHS). (Tribune Consortium, 1994, summarized from 97–129.)

TeleScopia

In the period between mid-1994 and mid-1995 the many telematics (tele-learning) demonstrator projects (as well as the more R&D-oriented projects) in the DELTA Programme and other programmes sponsored by the Commission of the European Community were coming to an end and a new round of project competition was underway for awarding later in 1995. To bridge this time period, the Commission stimulated with a sort of bridging function a special Joint Project initiative, in which four trans-European large-scale tele-learning demonstrator projects were underway (see Moeller, 1994 for a brief summary). One of these projects was *TeleScopia*, ('Trans-European Learning System for Cross Border Open and Interaction Applications'), a consortium of 13 project partners in eight European countries. In *TeleScopia*, six course providers are supported by (combinations of) three scenarios of telecommunications infrastructure – an interactive television network, ISDN video conferencing and videophony, and data communication via ISDN and PSDN – to deliver courses adapted for trans-European tele-learning. The courses are in different domains – managing innovation, environmental management, introduction to networks, English for specialized groups of professionals, financial controlling, and Greek as a foreign language. The target students include university students, practising engineers, those with financial responsibilities in small- and medium-sized businesses, and others, generally (but not entirely) members of professional groups. A scientific team from five countries is charged with the evaluation of the demonstrator project.

Evaluation Results from DELTA

The projects in the DELTA Programme completed in 1994 have been extensively evaluated, both vertically (within each Project, see for example, Sippel, 1994) and horizontally (across Projects). The interim results have been summarized by the ARTICULATE consortium and reported by the Tribune Consortium

(1994) and final results were to be submitted by the end of 1994. Thus the results and recommendations for a 'common evaluation framework' along with instruments to assist in the evaluation process will be presented to the field, but were not yet in convenient circulation for the initial planning of the Joint Projects such as *TeleScopia* which began before the DELTA Programme finished. The interim results of the horizontal evaluation, based not only on literature review and survey of practice in the field internationally, but also a mapping of the vertical evaluations within the component DELTA Projects on to various interpretive dimensions, are in circulation and have indicated, among other findings, that:

- Technology assessment in tele-learning projects should be embedded both within an environmental context and broad social and economic policy and issues. Thus evaluation methods and practices for tele-learning demonstrator projects must be highly sensitive to the (changing) questions asked by different stakeholders in the educational and training domains.
- There has been a marked shift toward accountability-oriented evaluation models, while at the same time, the 'positivist evaluation tradition, [including] control groups, has lost ground' (Tribune, 1994, 21).
- There is a necessity for evaluation projects to address cross-cultural issues.
- The problem of evaluating costs, especially in light of current demands for accountability to funders, is unresolved. (These findings are from the Tribune consortium summary, 1994, 20–21, which in turn was based on the ARTICULATE Project Deliverables, 1992a,b,c).

Thus evaluations for trans-national tele-learning demonstrator projects need to be context and policy sensitive, responsive (and thus adaptive) to questions of interest for their major stakeholders even as these emerge in the evaluation, and move away from 'significant results' design approaches toward approaches more sensitive to critical areas of interest to those supporting the demonstrator projects such as cost-related issues and language and cultural issues in trans-national settings.

CONCEPTUAL AND METHODOLOGICAL DIFFICULTIES

Despite initiatives such as that of the ARTICULATE Project, there are still fundamental difficulties in the

evaluation of trans-national tele-learning demonstrator projects, many of which can be traced to ongoing difficulties in the evaluation of the broader category of educational interventions of which the tele-learning demonstrator projects are a subset. Various problems are predictable, even without the added complexities related to breadth of scale, trans-cultural aspects and new technologies that characterize trans-national tele-learning demonstrator projects. Looking generally at 15 years of evaluation studies in educational contexts, and through a meta-analysis of 65 evaluations (defined as studies whose purpose was to communicate the results of a demonstrator project to a decision maker), Cousins and Leithwood (1986, 360) showed that the impact of this evaluation research depended on the extent to which:

- the findings were consistent with the beliefs and expectations of the potential users of the evaluation;
- those who were potential users of the evaluation had a prior belief in the value of the evaluation;
- potential users of the evaluation considered the data reported in the evaluation to be relevant to their problems;
- those involved in the evaluation felt the approach, intensity and methodological sophistication of the evaluation to be appropriate.

Such expectations are difficult to achieve, and, with regard to the first two points, may be outside the methodological control of the evaluator. The results of the meta-evaluation of Cousins and Leithwood further substantiated the impression that evaluation studies of demonstrator projects are characterized by a 'widespread non-use of results' (*ibid.*, 361). This conclusion came, despite the fact that a discipline of educational evaluation had been long established (more than three decades), along with an extensive body of theoretical literature and methodological guidelines (see for example, Worthen and Sanders, 1973, in their summary of already-in-1973 classic evaluation models for educational programmes and interventions).

Many have analysed the difficulties in carrying out effective evaluations of educational interventions involving new technologies. As a recent example, Draper and his colleagues (1994) reanalysed a broad set of evaluation studies and described a trend away from evaluating in the sense of making a value judgement about an intervention toward evaluating for a better 'discovery of how an intervention performs'; a trend toward the gathering of

information so that others can make such value judgements, rather than the evaluators making the judgements themselves. This in turn translates to gathering information likely to be relevant to those others who will make field-based decisions. Methodologically, it means:

not primarily focusing on theoretically showing that a given factor can affect learning in general, but rather on practically studying each situation with the question: 'What factors are most important in this teaching situation?' (*ibid.*, 8).

Draper and his colleagues further identified some of the many problems that confound the evaluation of educational innovations, including confounds related to student, instructor and organizational motivation; to various sorts of measurement problems; to variation in the situation of use of an innovation; and to the fundamental lack of an overall model of the teaching and learning process, despite many attempts at presenting such a model (*ibid.*, 9–14).

These and many other problems (see 'the gruesome dozen' cited by Achilles, 1982) continue to confound educational programme evaluation in general and the evaluation of educational innovations involving technology in particular (see Collis, 1989). Thus they will certainly trouble the evaluation of trans-national tele-learning demonstrator projects.

METHODOLOGICAL FRAMEWORK AND DESIGN FEATURES

Given these problems and the characteristics of trans-national tele-learning demonstrator projects, what are major aspects of a methodological framework and design features for the evaluation of such projects? This question needs to be addressed in terms of (a) a choice of a focus for the evaluation, based on the target audiences for the evaluation; (b) choice of a general model for the evaluation; (c) choice of a general strategy or method for the evaluation along with a consideration of some of the technical issues accompanying this choice; and (d) design decisions within the chosen method for the evaluation including planning for data capture and analysis within the chosen designs. Each of these aspects will be discussed separately.

Choice of a Focus for the Evaluation

Many different focuses have been suggested for evaluations (over 100 were identified by Patton in

1987), including focuses on: accessibility to participants, meeting of accreditation standards, appropriateness of services to client needs, cost-effectiveness, competencies demonstrated by participants, impact evaluation, longitudinal impact on participants, the strengths and weaknesses of programme delivery, reliability and replicability of results, and utilization-centered evaluation, among many. For a trans-national tele-learning demonstrator project, funded by a central agency whose aims are diffuse but include the desire to stimulate collaborative activity among different national partners and to contribute in time to the higher economic productivity and social harmony and equalization of the overall community (as are the goals of the Commission of the European Community for its member states, see Van den Brande, 1993) it seems advisable to choose a focus for the evaluation that is future-oriented around some aspect of the aims of the central agency rather than judgemental about the project components. A focus on how, and under what conditions, trans-national tele-learning may lead to more societal productivity and more social equalization, is such a choice. However, as the meta-evaluation of evaluations carried out by Cousins and Leithwood (1986) has shown, it is critical that all the stakeholders in an evaluation share a sense of the value of the evaluation in order for it to have eventual acceptance. Thus a focus on improving productivity would have to be seen as valuable not only to a central funding agency but also to key participants. In the case of trans-national telelearning, key participants could include educational institutions involved in the demonstrator projects, as well as technical providers, learning-materials developers, tutors and instructors, learners and the researchers themselves. While none of these stakeholders would be *per se* against the goal of improving productivity, the goal is perhaps too abstract and diffuse to motivate individual participation. Thus it may be useful to translate a broadly stated goal into an indicator of progress toward such a goal.

Flexibility as a focus

For example, productivity is frequently seen as related to increased flexibility in training possibilities. Increasing the flexibility of training offer is often seen as a way to increase the attractiveness and accessibility of training to the individual; to reduce costs for the supporting institution (as unnecessary or redundant training activities are avoided); to contribute to the increased participation of individuals

in training opportunities that might be inaccessible if the time, place, length, pace, and content (among other variables) of training are fixed but become more accessible when the learner has options for training participation. Thus increased flexibility could contribute overall to heightened system productivity from a macro-economic perspective (for an analysis of the assumption that more flexibility in training offer will lead to more productivity in overall training output, see Collis *et al.*, 1995, in a report summarizing a preliminary analysis conducted by the partners in the TeleScopia Project). Flexibility as a focus thus has the advantage of being an issue of immediate interest as well as intuitively attractive to the major players involved in a tele-learning demonstrator project. For the learner, which new options can be available to him in his study choices because of tele-learning? For the course provider, what must be done to adapt existing courses and course delivery in order to offer new options to a broader group of client learners? How can flexibility be brought into courses and pre-existing learning materials so that learners of different languages and cultures can participate? For those providing local support to learners involved in trans-national tele-learning, how can such support be managed, particularly as learners choose among many combinations of study pace, content, time, and other dimensions of flexibility? For decision makers, what are the financial implications of supporting institutional participation in more-flexible learning experience? Are the costs worth the payoff? For telecommunications providers, what flexibility options will increase the market for communications services in education and training?

Focus on a common goal for which something meaningful can be learned from the demonstrator project

Thus, in choosing an evaluation focus for a tele-learning demonstrator project whose roots lay in broad-scale initiatives for overall system productivity, it is strategic to choose to focus the evaluation around communal investigation of a common core issue of interest to all parties, rather than to focus the evaluation on judgement of the individual partners in their execution of the project.

With demonstrator projects involving innovative communication services, this seems particularly apt in that the current level of newness of such services to most participants is such that start-up problems in terms of the various stakeholders are predictable.

Thus a short-term demonstrator project should not spend its time arriving at conclusions that are already well-known: for example, that at this point in time, technical issues involved with new applications of telecommunications are still challenging; that learning materials must be redesigned to integrate in a meaningful way the possibilities made available by telecommunications; that organizations must support the innovative courses; that instructors and learners will sometimes resist change; that costs must be kept low (see Sippel, 1994, for an extensive summary). Nor can a short-term project expect to cause substantial change in institutional or system characteristics; a demonstrator project for tele-learning cannot in itself expect to result in impact on national economic indicators, (or even more far-fetched indicators such as reduction of traffic backups and even road fatalities in that the participants are not travelling to a traditional setting as much or at the same time as before. This presumption was used as the basis for a highly funded demonstrator project of tele-learning in The Netherlands, but given the small number of participants and a number of other factors the likelihood of such a study paying off in a measurable drop in highway facilities or reduction in traffic backups was unreasonable).

Choice of a General Model for the Evaluation

In their analysis of evaluation models for innovative educational programmes, Guba and Lincoln (1981) surveyed models dominating the 1940s to 1970s, and noted inadequacies in each of those models, especially with what might be called preordinate models. They summarize main characteristics of preordinate evaluations as: based on programme intentions and objectives; as having evaluation designs completed by the evaluator at the beginning of the evaluation; as relying on 'objective' data such as test scores; as providing feedback in formal reports, often only at the end of the evaluation; and as being based on paradigms from experimental psychology (*ibid.*, 28). In contrast, they summarize major characteristics of 'responsive models' (preordinate and responsive models have been defined and described by a number of evaluation theorists, initially by Stake, in 1975) as: being based on audience concerns, issues, and reactions and problems of persons participating in the entity being evaluated; as having continuously evolving evaluation designs; as relying on subjective data such as observations, interviews, and interactions; as pro-

viding feedback continuously as needed by audiences; and as being based on paradigms from anthropology and journalism (Guba and Lincoln, 1981, 28; Schermerhorn and Williams, 1979).

Given the findings of the ARTICULATE Project with respect to trans-European tele-learning demonstrator projects in DELTA, and other sources, the following conclusion had been earlier drawn with respect to the evaluation of such projects:

evaluations for trans-national tele-learning demonstrator projects need to be context and policy sensitive, responsive (and thus adaptive) to questions of interest for their major stakeholders even as these emerge in the evaluation, and move away from 'significant results' design approaches toward approaches more sensitive to critical areas of interest to those supporting the demonstrator projects such as cost-related issues and language and cultural issues in trans-national settings.

The key aspects of this conclusion map directly on the key characteristics of responsive models for evaluation, rather than preordinate models. Thus it seems reasonable to choose a responsive model, with the following main aspects, for evaluation of tele-learning demonstrator projects:

- responsive to audience concerns rather than to investigating if preset objectives are being met;
- having a continuously evolving evaluation design, rather than a fixed set of procedures at the start of the evaluation;
- relying heavily on subjective data such as interviews, observations, and other forms of interactions with participants and stakeholders.

Choice of a General Strategy for the Evaluation

Having chosen to focus the evaluation on a focus question of interest to all stakeholders, and to adopt a responsive approach, the next decision in the evaluation planning is the choice of a general strategy or method for the evaluation. Major methods for evaluation research include: experimental methods, involving the controlled comparison of a treatment group with a control group; evaluation based on grounded theory; survey methods; archival or historical analysis; and case studies (Yin, 1989, 17). A case study is defined as an empirical inquiry that: 'investigates a contemporary phenomena within its real-life context, when the boundaries between phenomenon and context are not clearly distinct, and in which multiple sources of evidence are used' (*ibid.*, 23).

Yin further defines a case study methodology as most appropriate for research questions starting with 'How?' or 'Why?' which are being asked about a contemporary set of events, over which the investigator has little or no control (20). Also 'What?' questions could be added to this list. In a tele-learning demonstrator project, where a unifying focus is on a question such as 'How can training be adapted to become more flexible?', where the context is a contemporary set of events in terms of the activities happening in the demonstrator project, and where the evaluators are not designating what and how the demonstrator-project activities should be executed, the choice of a case study methodology seems appropriate.

Much has been written about case study methods (Yin, 1989, 1993, is a major source) but even so there are many misconceptions. There are those who think case study methods are synonymous with lack of vigour, that one can just 'go out and do a case study'. Another complaint is the possible lack of generalization, from one case to a population. A third fear is that a case study will generate a massive amount of data and documentation which will in turn take much time to analyse. Yin (1989) refutes all of these concerns, if case studies are properly carried out.

Given the concern relating to lack of vigour in case studies, it is unfortunate that no 'catalogue' of research designs for case studies has been generally accepted. According to Yin (1989) there are no standard textbooks dealing with such issues as selection of subjects, the selection of various stimulus conditions, and the identification of various response measures, as there are for experimental and quasi-experimental design. Nor are there common design approaches such as have emerged for survey research. For a considerable time, the case study was equated with 'one-shot, post-test only' quasi-experimental designs, felt weak in both validity and reliability, an interpretation that has been refuted (Cook and Campbell, 1979). In an appropriately carried out case study, it is critical that a study's questions are clear; that there are some initial propositions, not hypotheses but initial ideas related to the answer to the study's questions; that the unit of analysis is clear to all involved and relates to the research questions; that there is a strategy for linking data to propositions, such as pattern matching and replication in multiple-

case studies; and that there are criteria for interpreting the findings of the case studies (Yin, 1989, 27-60). Thus replication and multiple-case designs, instead of sampling logic, forms the basis for the selection of cases. Cases may be chosen where similar results are expected or where contrary results are expected. A type of flexibility offered by case study designs is the possibility throughout the research to select other cases different from those initially identified, if these contribute to the better investigation of the research question.

Thus for trans-national tele-learning demonstrator projects the choice of a multiple-case study design seems well indicated, if accompanied by a cohesive research question or focus, the articulation of initial propositions to steer data collection, the designation of the unit of analysis, a strategy to link propositions to observations, and criteria for interpreting the results.

Design Decisions

Once the decision for a multiple-case study method has been made, decisions must be made relative to data capture and analysis. Overriding these is another conceptual decision: the determination to use a mixed-method evaluation design. Mixed-methods are sometimes called triangulation, the designed use of multiple methods, with offsetting or counteracting biases, in investigations of the same phenomenon:

The core premise of triangulation as a design strategy is that all methods have inherent biases and limitations, so use of only one method to assess a given phenomenon will inevitably yield biased and limited results (Greene *et al.*, 1989, 256).

There are other approaches to mixed methods besides triangulation, including complementarity; developmental approaches, where results from one method are then used with another method; and expansion, where different methods are used for different inquiry components. In practice, these result in the expectation that more than one method of collecting data will be used in a case study.

These methods and analysis strategies differ relative to whether the data collection is seen as within the individual case, with a set of multiple cases running in parallel, or if the multiple-case study also involves the horizontal examination of the case studies as a set, in terms of a particular research questions. Both approaches are desirable.

Techniques for with-case data handling

When conducting a within-case case study, there are generally six sources of evidence (Yin, 1989). These include: documentation (letters, agendas, proposals, progress reports, newsclippings, etc.); archival records (organization charts, service records such as number of persons served in a given time, lists of names, survey data previously collected; personal records such as diaries, calendars and traces of telephone calls or keystrokes from use of a computer program); interviews, open-ended, focused, or survey-type; direct observations; participant observation, where the observer is also taking a role within the case being studied; and physical artifacts, tools or instruments or other physical evidence of the process or product being carried out within the case. Also according to Yin, there are three main principles to data collection: using multiple sources of evidence; creating a case study data base; and maintaining a chain of evidence, whose principle is to allow the eventual reader of the case study a way to follow the derivation of conclusions about the case from initial research questions through the data to the conclusions.

The general analytic strategies for within-case data differ if the researcher is predominately exploring and describing or explaining and predicting (Miles and Huberman, 1994). For exploring and describing, there are techniques such as context charts and checklist matrices; time-ordered displays, such as event listing, critical incident charts, activity records, and time-ordered matrices. There can also be role-ordered displays, role-by-time matrices, and conceptually ordered displays including conceptually clustered matrices and cognitive maps. Within-case displays used to support predictions are more stringent in their requirements than those used to describe and explore. Techniques such as causal networks and explanatory effects matrices can be used. The classic reference for techniques for displaying and interpreting qualitative data is Miles and Huberman, (1984, 1994), complete with hundreds of graphic and matrix-based examples.

Techniques for cross-case data handling

Similarly, Miles and Huberman present techniques for dealing with cross-case displays, resulting from the study of a set of cases in order to address key questions running through all cases. There are techniques for displays for simultaneously exploring and describing multiple cases, such as conceptually ordered displays, case-ordered displays, and time-

ordered displays. There are also techniques for cross-case analysis whose purpose is to explain and order, such as causal models, variable by variable matrices, and case-ordered effects matrices (see Miles and Huberman, 1994, 172–238).

Drawing conclusions

Drawing and verifying conclusions from any of the data display and reduction techniques requires (a) tactics for generating meaning, such as noting patterns, themes, seeing plausibility, clustering, partitioning variables, and finding intervening variables; (b) tactics for testing or confirming findings, such as checking for researcher effects, triangulating, checking the meaning of outliers, ruling out spurious relations, and looking for negative evidence; and (c) standards for the quality of conclusions, based on criteria such as confirmability, authenticity, and transferability (Miles and Huberman, 1994, 245–87). An important point in all these techniques is that choices for display technique should emerge throughout the case studies, not be predetermined. Based on the nature of the data and the patterns and counter-patterns emerging from them, the researchers should in an on-going manner investigate different sorts of display techniques.

Thus the conclusions for case study evaluations of trans-national tele-learning demonstrator projects are (a) carry out both within and across-case analysis, (b) use a variety of sources of evidence, (c) continually explore different display techniques for data, (d) give particular attention to tactics for generating meaning, for testing or confirming findings, and for supporting the quality of conclusions.

THE TELESCOPIA PROJECT

Earlier it was noted that the TeleScopia Project could be classified as a trans-national tele-learning demonstrator project and its particular characteristics were noted. One of the three components of TeleScopia is its research and evaluation component (the other two relating to course providers and telecommunications providers). The research-component group has been interacting for two months on its conceptualization of a methodological framework and design features for its task of evaluation within TeleScopia. This paper so far has been a summary of the theoretical outcome of the interaction, and although written by one of the partners reflects interaction among all the partners

involved in the research component. Applying the principles, the following plan for the evaluation has emerged (its more specific details appear in an appendix, for internal use only).

Choice of Flexibility as a Focus for the Evaluation

As explained in the document 'Flexibility as a Key Construct in European Training: The TeleScopia Project' (Collis *et al.*, 1995) the research team has taken as an assumption that a movement from training offer fixed in various of its key dimensions toward training offer in which the learner can make choices from a number of options will eventually lead to more productivity in training output. From this assumption, a general focus for the evaluation in the TeleScopia project has become: 'What can we learn from the TeleScopia courses about movement from a traditional training paradigm (which has most of its aspects fixed from the point of view of the learner) to a future training paradigm emphasizing more flexibility in the options open to the learner?'

As explained above, flexibility is seen as important to each of the component groups in TeleScopia, to the course providers, to the technology providers, and to the researchers, as well as to the funder of the project, the Commission of the European Community.

From this focus on flexibility, a set of three key research questions has emerged:

1. What dimensions are most important with regard to the FIXED to FLEXIBLE transition for European training?
2. What factors related to tutors, course-providing agencies, learners, learners' employers, technologies, instructional design, learning materials, costs, culture, social and legal issues most critically constrain or stimulate the movement from FIXED to FLEXIBLE on these key dimensions?
3. How can the different contributions of telematics, particularly through the metaphors of 'the virtual classroom', 'extended contacts', and 'the collaborative learning group' most powerfully contribute to the movement from FIXED to FLEXIBLE learning?

And from these, a subsequent set of subquestions is being addressed (numbered sequentially for convenience):

1. How can the differences among cultures, including language differences, be most effectively dealt

with when adapting learning materials developed elsewhere for local practice when participating in trans-European tele-learning?

2. What are central guidelines for adapting existing courses for a more-flexible future involving trans-European tele-learning?
3. How can different varieties of instructional organization, including self-study and cooperative learning, be made available within courses for trans-European tele-learning?
4. How can instructional instrumentation such as groupware and simulations be integrated into trans-European tele-learning?
5. What is the relationship between the technical platforms involved in trans-European tele-learning and the instructional methods and materials that are employed?
6. How can support for learners participating in trans-European tele-learning be best organized?
7. How can local tutoring be best organized for trans-European telelearning?
8. In moving from a fixed to a more-flexible training situation, what are the cost implications, in what direction, and for whom?

These questions reflect a major point of Draper's (1994) analysis: a movement in evaluation away from questions that ask 'Does it cause learning?', toward a consideration of factors that are most important in a learning context. The questions also reflect the Tribune consortium's finding that cross-cultural issues and cost issues are of critical importance in tele-learning, both in demonstrator projects and long-term. The subquestions emerged from various discussions with the course providers, preliminary visits to some of the course-provider sites, and different forms of interaction among the research team itself, thus reflecting the principle of allowing questions to emerge from the stakeholders in an evaluation.

Thus, the TeleScopia research is focused on an issue generalizable outside the project itself, and of strategic interest to its participants and its funding agency.

Choice of a Responsive Model for the Evaluation

The TeleScopia has chosen a responsive, rather than a preordinate model for the evaluation. In particular, this means there is an intention to remain alert throughout the project to audience concerns and a mechanism for systematic integration of those concerns into the rest of the evaluation (this is

explained further in this paper). In order to stimulate a responsive approach to the evaluation, a basic communal strategy has been put forward for a first phase of the evaluation, and provision has also been made for the research team to come together regularly, and assess the value of integrating new audience concerns into their overall evaluation. Four such communal meetings are planned. In each of these there will be provision for the evaluation partners to compare their impressions of emerging audience concerns and to integrate those of sufficient impact to the project into the planning of the evaluation or the remainder of the project. Also, following the lines of the responsive evaluation model (Stake, 1975; Guba and Lincoln, 1981) predominately subjective data, from as many participants as relevant, rather than test-score oriented objective data from the learners only, will be collected.

Choice of the Case Study Method for the Evaluation

Within its chosen responsive model of evaluation the TeleScopia Project has chosen a multiple-case study method to address its research questions. Each of the TeleScopia courses being adapted for trans-European tele-learning within the project will serve as a unit of analysis. A particular evaluation team from the TeleScopia research partners will be responsible

throughout for each of the courses, carrying out a within case case study of that course. Figure 1 shows this in the form of six vertical columns, each going on in time throughout the project. Each column represents one of the courses in TeleScopia and thus the focus of one of the case studies.

The multiple cases will also be used for triangulation and complementarity with each other; thus each case will include some of the same sources of information collected in the same way as in each of the other courses-as-cases. Individual cases may also choose to add some other sources of data; particularly promising data emerging from such an addition will be shared with the other case-study researches, through discussion during regular meetings of the evaluation team, with the opportunity to integrate the new data-source technique into all the within-case studies. This responsive process will occur during four 'cross-case' meetings of the evaluation team.

A key feature of the multiple-case study method is that it allows the execution of parallel within-case case studies and also the execution of cross-case analysis through addressing the project research questions through the synthesis of experience in all the cases. Figure 2 adds the cross-case aspect to the within-case case-study aspect of the evaluation shown in Figure 1:

Figure 1. The six case studies in the TeleScopia project

| Time Periods: | Case Study 1: Berlitz | CS2: La Sept/Arte | CS3: Enie | CS4: UETP Macedonia | CS5: UETP EEE | CS6: IWB |
|---------------|-----------------------|-------------------|-----------|---------------------|---------------|----------|
| Jan. 95 ↓ | | | | | | |
| Nov 95 | Final Report | | | | | |

Figure 2. Cross-case research questions

| | CS1: Berlitz | CS2: La Sept/Arte | CS3: Enie | CS4: UETP Macedonia | CS5: UETP EEE | CS6: IWB |
|-----------------------|--------------|-------------------|-----------|---------------------|---------------|----------|
| Phase 1 | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ |
| Cross-Case Meeting 1: | | | | | | |
| Phase 2 | | | | | | |
| Cross-Case Meeting 2: | | | | | | |
| Phase 3 | | | | | | |
| Cross-Case Meeting 3: | | | | | | |
| Phase 4 | | | | | | |
| Cross-Case Meeting 4: | | | | | | |
| Phase 5 | | | | | | |
| | Final Report | | | | | |

Important in this figure are: (a) the six case studies, based on the courses, being carried out throughout the project with each research team taking responsibility for one or two of the case studies and each involving a set of common sources of data and instrumentation as well as being open to other types of data collection in the individual case study; (b) the on-going consideration of the eight research subquestions by each of the research teams throughout the project; thus each researcher can choose the research question of most interest to him or her and be responsible for its consideration throughout. This consideration will occur within the researchers own case study, but also across cases, through regular provision for the researchers to meet together and share data and experiences; and (c) the periods of cross-case discussion and analysis, involving all the researchers will result not only in new insights into the research questions, but also possibly into a revision of some of the questions and a revision of the data-collection and analysis procedures within the cases during the next phase of the project, until the researchers come together again.

Phase 1, from January to February 1995, consisted of baseline data collection in each of the cases, along with each researcher being alert to ideas as to how to investigate his or her particular research question both within his or her case and also across cases. The first cross-case meeting involved a communal discussion of the researchers' experiences with the parallel data collection within the cases, an assessment of possible adaptations of some of those techniques for data collection, an exploration of ways to display the data captured so far in the cases, and first discussions of the research questions from a cross-case perspective. The remaining phases repeated this cycle with interspersed communal sessions.

The final report was delivered in November 1995.

Design Decisions Within the Case-Study Method

During the first phase of the evaluation, the researchers will predominately be getting preliminary information on their particular cases. For this each researcher will be asked to assemble for his case:

- Documentation: letters, memoranda, agendas, announcements and minutes of meetings relating to the course, other written reports of events, newsclippings, administrative documents such as project proposals.
- Archival records, including a record of the number

of learners taking the courses previously and during TeleScopia, organizational charts and records about the course providers relevant to the courses, maps and charts of the geographic characteristics of the course-provider's work setting, and of the locations of sites for the courses, photos or video if available, lists of names of all relevant persons in the course, diaries, logbooks, printouts of computer conferences, and other archival information.

- A first plan for who will be interviewed within each course, during the period February–April.
- A first plan of how often, and where, direct observation will take place in the course, and perhaps at a learner site, during the period February–April.
- A cohesive overview of the course materials for each course, before adaptation for TeleScopia and a clear indication of what aspects of the course materials are being adapted for TeleScopia and a description of what aspects of student work can be saved for further analysis.

Figure 3 shows the matrix in Figure 2, expanded per cell for the first of the evaluation phases.

| Questions, Across all Cases | Techniques within each case | | | | | | | | |
|--|-----------------------------|-----------------|----------------------|---------------------------|-----------------------------|------------------|--------------------|----------------|------------------------|
| | Observations | Screen displays | Videotape of context | Print-out from comp. conf | Technical admin. interviews | Tutor interviews | Learner interviews | Anecdotal logs | Calendar of activities |
| 1. How can the differences among cultures, including language differences, be most effectively dealt with when adapting learning materials developed elsewhere for local practice when participating in trans-European telelearning? | | | | | | | | | |
| 2. What are central guidelines for adapting existing courses for a more-flexible future involving trans-European telelearning? | | | | | | | | | |
| 3. How can different varieties of instructional organization, including self-study and cooperative learning, be made available within courses for trans-European tele-learning? | | | | | | | | | |
| 4. How can instructional instrumentation such as groupware and simulations be integrated into trans-European telelearning? | | | | | | | | | |
| 5. What is the relationship between the technical platforms involved in trans-European telelearning and the instructional methods and materials that are employed? | | | | | | | | | |
| 6. How can support for learners participating in trans-European telelearning be best organized? | | | | | | | | | |
| 7. How can local tutoring be best organized for trans-European telelearning? | | | | | | | | | |
| 8. In moving from a fixed to a more-flexible training situation, what are the cost implications, in what direction, and to whom? | | | | | | | | | |

Figure 3. The first evaluation phase

Within each phase of the evaluation, the set of data-gathering techniques initially including those in the columns will be used within each case study, in particular to collect information relevant to the research questions shown in the rows of the figure. In addition, each researcher will consider his or her own cross-case research question, and bring to the first cross-case meeting requests for how questions or probes about his question could be integrated into the interviews and observations per site.

Among the tasks of the first cross-case meeting will be the planning for the setting up and maintenance of a data base for the cases. Another task will be a first selection of documents, archival objects, and physical artifacts of the project for common analysis and summary and a selection of various data-display techniques to help visualize trends or patterns in the data. A further task will be the communal development of interview protocols and observation plans in order to address the research questions of each of the researchers. Finally, an overall calendar will be made of when site visits will be made and when and how data flow should occur among the cases during the phases of within-case activity.

CONCLUSION

There are a number of innovative aspects in this framework. The idea of each researcher focusing on his or her own research question across the cases while also being responsible for a within-case study itself is perhaps innovative. The regular provision for responsiveness and for systematic integration of suggestions from this responsiveness into the overall evaluation plan should stimulate growth of ideas within the project. The iterative aspect also relieves the load of before-the-fact planning of the details of an evaluation data-collection scheme, details which can often underestimate the burden of the evaluation on the participants or require eventual adjustment for natural changes in the project. With each researcher having clear responsibilities within cases and across cases, the management of evaluation activity may be clarified. The likelihood of results still relating to the original research questions is increased, as those questions will be continually someone's focus. The participants in the rest of TeleScopia will have to evaluate the evaluation as to its intensity and methodological sophistication.

REFERENCES

- Achilles, C M (1982) The gruesome dozen: Problems plaguing program/project evaluators, *Educational Evaluation and Policy Analysis*, 4, 4, 439-42.
- ARTICULATE, (1992a) *Preliminary Common Evaluation Framework*, ARTICULATE Deliverable No 1, Commission of the European Community, Brussels: DG XIII.
- ARTICULATE Deliverable No 2, Commission of the European Community, Brussels: DG XIII.
- ARTICULATE, (1992b) *DELTA Pilots and Development Projects and their Evaluation*.
- ARTICULATE, (1992c) *The Evaluation Help Desk*, ARTICULATE Deliverable No 6, Commission of the European Community, Brussels: DG XIII.
- Berk, R A and Rossi, P H (1990) *Thinking About Program Evaluation*, Sage, Newbury Park, CA.
- Collis, B (1989, March) *How should we evaluate 'dynamics and vision'?*, Study for the International Customer Executive Seminar, IBM International Education Center, La Hulpe, Belgium.
- Collis, B, Vingerhouts, J and Moonen, J (1995) *Flexibility as a key construct in European training: The TeleScopia Project*. The paper was a deliverable of the TeleScopia project funded by the DELTA Programme of the DG XIII of the Commission of the European Community.
- Cook, T D and Campbell, D T (1979) *Quasi-experimentation: Design and Analysis Issues for Field Settings*, Rand McNally, Chicago.
- Cousins, J B and Leithwood, K A (1986) Current empirical research on evaluation utilization, *Review of Educational Research*, 56, 3, 331-64.
- Deutsche Bundespost Telekom (1994) Project Description, TeleScopia, Generaldirektion, Bonn.
- Draper, S W, Brown, M I, Edgerton, E, et al. (1994) *Observing and measuring the performance of educational technology*, A report by the University of Glasgow institutional project in the Teaching and Learning Technology Programme (TLTP) University of Glasgow, Robert Clark Centre for Educational Technology, Glasgow.
- Glennan, T K (1987) *The Role of Demonstrations in Federal R&D policy*, The Rand Corporation, Santa Monica, CA.
- Greene, J C, Caracelli, V J and Graham, W F (1989) Toward a conceptual framework for mixed method evaluation designs, *Educational Evaluation and Policy Analysis*, 11, 3, 255-74.
- Guba, E G and Lincoln, Y S (1981) *Effective Evaluation*, Jossey-Bass Publishers, San Francisco.
- Madaus, G F, Stufflebeam, D and Scriven, M (1989) Program evaluation: A historical overview, in Madaus, G F, et al. (eds) *Evaluation models: Viewpoints in Educational and Human Services Evaluation*, Kluwer-Nijhoff, Norwell, MA.
- Miles, M B and Huberman, A M (1984, 1994) *Qualitative data analysis*, First and second editions, Sage, Beverly Hills, CA.

- Moeller, M (1994) The new telematics programme: Shaping the information society, *JANUS User Association Pilot Issue*, September, 7.
- Patton, M Q (1987) *Creative evaluation*, Sage, Beverly Hills, CA.
- Schermerhorn, G R and Williams R G (1979) An empirical comparison of responsive and preordinate approaches to program evaluation, *Education Evaluation and Policy Analysis*, 1, 3, 5560.
- Sippel, F (ed) (1994) *Implementation and management of teletraining*, Proceedings of a Workshop held in Brussels, 30 November, 1993. The workshop was a deliverable of the Multimedia Teleschool (MTS) Project (D 2021) funded by the DELTA Programme of the DG XIII of the Commission of the European Community.
- Stake, R (ed) (1975) *Evaluating the Arts in Education: A Responsive Approach*, Merrill, Columbus, OH.
- TRIBUNE Consortium (1994) *DELTA at Work: A snapshot*, Tribune European Learning Technology Collection, No 3. Bollman-Druck GmbH, Zirndorf, Germany.
- Van den Brande, L (1993) *Flexible and distance learning*, John Wiley, Chichester.
- Worthen, B R and Sanders, J R (eds) (1973) *Educational Evaluation: Theory and practice*, Wadsworth Publishing, Belmont, CA.
- Yin, R K (1989) *Case Study Research: Design and Methods*, Sage, Beverly Hills, CA.
- Yin, R K (1993) *Applications of Case Study Research*, Sage, Beverly Hills, CA.

BIOGRAPHICAL NOTES

Betty Collis and **Jolande Vingerhoets** both work in the Faculty of Educational Science and Technology at the University of Twente, The Netherlands.

Address for correspondence: Faculty of Educational Science and Technology, University of Twente, PO Box 217, 7500 AE Enschede, The Netherlands.

STAFF AND EDUCATIONAL DEVELOPMENT SERIES

Published in association with the Staff and Educational Development Association

Series Editor: Sally Brown, Educational Development Advisor; University of Northumbria at Newcastle

- staff development in higher and further education
- practical innovation in teaching and learning
- best practice, theory and debate

NEW

Enabling Student Learning Systems and Strategies

Edited by Gina Wisker and Sally Brown
£18.99 Paperback ISBN 0 7494 1790 0 192 pages
March 1996 Order No KS 790

Assessing Competence in Higher Education

Edited by Anne Edwards and Peter Knight
£18.95 Paperback ISBN 0 7494 1458 8 192 pages
1995 Order No KS 458

Assessment for Learning in Higher Education

Edited by Peter Knight
£18.95 Paperback ISBN 0 7494 1532 0 192 pages
1995 Order No KS 532

Research, Teaching and Learning in Higher Education

Edited by Brenda Smith and Sally Brown
£18.95 Paperback ISBN 0 7494 1412 X 224 pages
1995 Order No KS 412

Order now from your bookseller or direct from Kogan Page (please add 10% p&p):

120 Pentonville Road, London N1 9JN
Tel + 44 (0) 171 278 0433
Fax + 44 (0) 171 837 6348