Explorations in graphical argumentation

The use of external representations of argumentation in collaborative problem solving

PROEFSCHRIFT

Ter verkrijging van de graad van doctor aan de Open Universiteit Nederland op gezag van de rector magnificus prof. dr. ir. F. Mulder ten overstaan van een door het College voor Promoties ingestelde commissie in het openbaar te verdedigen op vrijdag 20 juni 2003 om 16.00 uur precies

door

Jan Marinus van Bruggen
geboren op 23 maart 1953 te Holten

Promotores:
Prof. dr. W.M.G. Jochems, Open Universiteit Nederland
Prof. dr. P.A. Kirschner, Open universiteit Nederland

Overige leden van de beoordelingscommissie
Prof. dr. H.M.J. van den Bosch, Open Universiteit Nederland
Prof. dr. H.P.A. Boshuizen, Open Universiteit Nederland
Prof. dr. A.J.M. de Jong, Universiteit Twente
Prof. dr. G. Kanselaar, Universiteit van Utrecht
Prof. dr. F.L. Leeuw, Universiteit van Utrecht

www.ou.nl/otecresearch

Omslag: Open Universiteit Nederland

Druk: Datawyse Maastricht

Hence, to the extent that the mind is unable to juxtapose consciously a larger number of ideas, each coherent group of detailed constituent ideas must be reduced in consciousness to a single idea; until at last the mind can consciously juxtapose them with due attention to each, so as to produce its single final idea.

From: Wigmore, J.H., *Principles of judicial proof or the process of proof as given by logic, psychology, and general experience and illustrated in judicial trials (2nd ed.* . Boston: MA, Little, Brown and Co, 1931.

Preface

The citation of Wigmore, which serves as a motto for this work, stems from a period in which logic, psychology and human experience could be referred to as equivalent sources for legal proof. To help overcome the limits of cognitive capacity Wigmore presents a notation for the graphical representation of legal argumentation in proof. Today we would call such a tool a 'cognitive tool'.

The psychological and educational interest in graphical representations of conceptual structures such as argumentation, seem to follow a cyclical pattern. About twenty years ago, 'spatial learning strategies', such as concept mapping or schematizing, attracted the attention of several researchers. During that period I was a member of the COWO team that was further developing and evaluating the techniques of schematizing, a method to graphically represent the concept structures in study texts. A few years later the research community lost its interest in these representations, until they made their comeback as 'cognitive tools' or 'mindtools'.

Many persons have contributed to the realization of this PhD thesis, in particular my promotores, Wim Jochems and Paul Kirschner. Wim Jochems expressed his confidence in this doctorate project, without batting an eyelid, and without having any description of the subject available. Dear Wim, I thank you for always keeping up your support even when the progress of the project did not seem to warrant it. Dear Paul, you offered to be my daily advisor without knowing the research topic or realizing what you were doing to yourself. Sorry for the latter. We shared years of interesting research, laughs, personal sadness and … Fawlty Towers. I had never before enjoyed the company of a guest that orders a pair of pliers!

Dear Otec colleagues, forgive me for not mentioning all of you here by name. You created the supportive environment needed to complete these type of projects. Wim van der Vegt and Wim Slot provided technical support: the problems reported in chapter 3 had nothing to do with your fine work. In the completion phase of this thesis the Otec secretaries, Marina Pongraz, Alice Boereboom-Pierey and in particular Ingrid Jonkman, provided important assistance in the type-writing. Thanks for that! Data scoring and coding were done by Andrea Rau, Femke Kirschner and Marijn Bruinink and I hope I can count on you for future projects.

Finally, dear spouse and children - Diny, Carine, Guido and Leonie – you have been the main victims of this project. I'm back!

Contents

1 General introduction

The Open University of the Netherlands is aiming to realize competency-based education with its students mastering complex skills. To realize this goal, students need to be confronted with *authentic* tasks that closely mimic those encountered in professional settings. A foundation claim of this thesis is that 'social science problems' (Voss, Tyler, & Yengo, 1983) or 'wicked problems' (Kunz & Rittel, 1970; Rittel & Webber, 1984) are excellent examples of these types of tasks. Examples of this type of problem are 'How can we reduce traffic fatalities?', 'How can we reduce aggression in commuter traffic?', 'How can we reduce school drop-out?', 'How can we stimulate self-regulated learning?' or the now classic problem 'How can we increase Soviet crop production?'.

Solving social science problems is often the task of (multidisciplinary) teams, which in the context of higher distance education will mean that we make use of collaborative learning settings. It is here where the student population of the Open University of the Netherlands offers interesting, though challenging, opportunities. Many, but certainly not all, of our students are already occupied in jobs in or closely related to the domain of their study. Therefore, although they may not yet operate on the level of (academic) domain experts, they certainly cannot be compared to novices in the domain. Other students of the Open University of the Netherlands, however, have no such domain experience. They are indeed novices in the domain. These different students may find themselves in the same team that tries to solve a social science problem.

The initial question of the work reported here was how external representations could be used to facilitate collaborative solving of ill-structured problems in computer-supported asynchronous settings. This is a rather broad question, and we therefore have to clarify and specify several of the elements contained in it, in particular the:

1. type of ill-structured problem to be supported
2. content and form of the external representations: what is represented and how it is represented
3. role that external representations play in the problem solving process
4. intended audience – who the problem solvers are that will use these external representations, and the context in which they will use them
5. extent to which we design and implement a computer-based environment to support problem solving with external representations.

First¸ the type of ill-structured problem needs to be clarified. To be more specific, the kind of ill-structured problems that we are dealing with have been referred to as social science problems (Voss, Greene, Post, & Penner, 1983; Voss, Tylor, et al., 1983). The classic example problem used by Voss is that of how to increase the crop production in the Soviet Union. Voss, Tyler, et al. describe the characteristic problem solving strategy for social science problems as "Identify the cause(s) of the problem and solve by eliminating the cause(s)" (Voss, Tylor, et al., 1983, p. 208). They found that a typical expert solution consisted of a few rather abstract solutions, whereas novices tried to isolate different causes and defined solutions in terms of eliminating these individual causes.

Another characteristic of social science problems is that one cannot be certain beforehand that a solution will work, because it may take years before a solution is implemented and produces observable results. Voss, Tylor, et al. (1983) use the concept *delayed evaluation* to describe this characteristic which prohibits the solver of social science problems to *prove* that a solution will work. The problem solver can only provide arguments as to why a solution can be expected to work, or why one solution is more likely to work than another and thus should be preferred over another. In other words, solving, or rather evaluating the solution of social science problems requires argumentation and since there is no way to provide formal proof that a solution will work, one needs to rely on informal reasoning.

Elsewhere, this kind of problem has been called ‘wicked problems’ (Kunz & Rittel, 1970; Rittel & Webber, 1984). The descriptions of social science problems and wicked problems overlap to a large extent, but there are some differences in accent. Research in social science problem solving has typically used single agent settings. In contrast, wicked problems are seen as problems that always involving many agents, or stakeholders, each of whom may bring a particular viewpoint on the problem. It makes no sense to study wicked problems using single agent settings. In chapter 2 we give a more detailed overview of these two approaches.

More recent psychological research has studied the way in which *teams* of problem solvers coordinate their problem solving activities involves sharing of data and/or problem solving operators (Boshuizen & Tabachneck-Schijf, 1998). As an example of these coordination mechanisms, Alpay, Giboin, and Dieng (1998) describe the use of several (shared) representations in multi-disciplinary accidentology teams. We will demonstrate in chapter 2 that by using a multiple agent, multiple representation perspective, the approaches of social science problem solving and wicked problem solving can be merged.

Second, we have to clarify what is meant by external representations that support the problem solving activities. In particular, we have to be clear as to *what* is represented in these external representations and in *what way* (visual, auditory…) it is represented. Here, w*hat* is represented is the *argumentation* related to the analysis and solution of the problem. As stated above, when one solves a social science problem one has to deal with *delayed evaluation* and this is achieved by supplying arguments as to why a particular solution should be preferred above another (it has worked in the past, it is similar to a known solution, et cetera). This argumentation is used to strengthen a solution. Argumentation can also play a role in problem solving activities. In their discussion of wicked problems Kunz and Rittel (1970) describe the entire problem solving process as an issue-based argumentative activity.

Given the core importance of argumentation to the process of solving these problems, the support that we try to offer is primarily aimed at the representation of the argumentation underlying the solution of this kind of problem. We will, however, not limit the support to the argumentation dealing with the delayed evaluation of social science problems. From the perspective of wicked problems both analysis and solving a problem need argumentation and we intend to use external representations of argumentation for these activities as well.

Using external representations to represent *argumentation* sets the work aside from an important line of research dealing with solving problems using (multiple) representations, such as simulations, modeling, diagrammatic representations, or charts (Van Someren, Reimann, Boshuizen, & de Jong, 1998). We are not dealing with those kinds of representations at all. For the time being - more accurate terms to describe various characteristics of external representations are introduced in chapter 2 – we describe the way’ in which we represent argumentation as being *graphical*, using sets of symbols that denote the objects and relations in the argumentation. Other terms to describe this use of external representations are 'graphical argumentation' (Buckingham Shum, MacLean, Bellotti, & Hammond, 1997) or 'argumentation visualization' (Kirschner, Buckingham Shum, & Carr, 2003), that is, the expression (by the users or a technical mediator) of an argumentation with the use of graphical symbols.

If one intends to use an external graphical representation of argumentation, one has to decide on the particular objects and relations that are available for use. The objects and relations and the rules for combining them are referred to as the 'representational notation' (Suthers, 2001; Suthers, in press). Each representational notation is biased, in other words one notation allows expression of some aspects of reality better or more smoothly than others. Depending on what needs to be expressed (by whom) a general set of objects and relations may suffice. Such a general set can be as simple as allowing the user to express a claim and its supporting or contradicting reasons, as is the case in ReasonAble® (Van Gelder, 2003). An example of a general set that is more complex is a Toulmin structure (Toulmin, 1958), where warrants and backings can be used to add support to a claim and qualifiers and rebuttals are available to narrow down the claim. One may also use domain-dependent sets, such as Decision Representation Language that was defined to support qualitative decision making (Lee & Lai, 1991). There are many trade-offs to be considered here, but we will limit ourselves to a single one here. Although domain-dependent representational notations will allow, in principle, the most accurate representation, these notations need to be mastered before they can be applied smoothly. Thus, one of the most important questions that need to answered is *what* the representational notation of the external representations must contain. What are the objects and relations in the notation? What, if any, are the rules about how they can be combined? The question here is not what is needed to (*formally*) represent argumentation, but rather how to define a representational notation that is workable – close enough to what problem solvers already do - to make it easier to include in their activities and adding new elements that add to their activities.

Third, one may ask whether there are any indications that external representations can play a support role in collaborative problem solving and if so, what that role is. There are indeed areas that indicate that problem solvers can benefit from using external representations. The first of these areas is that of computer-supported collaborative learning (CSCL). Research on CSCL environments for scientific inquiry provides evidence that the use of external representations of argumentation helps learners to maintain focus in problem solving (Suthers, in press; Veerman, 2000). Maintaining focus is identified as one of the success factors in collaborative problem solving (Erkens, 1997). The research in CSCL is limited in that it only addresses scientific questions, or the *why-part* of problem solving. A second area that offers relevant insights is aiming more at the *how-part* of problem solving in areas, such as human-computer interaction. Work on design rationale provides examples of how design argumentation can be modeled and stored for re-use. Much of the research and system development in this area is influenced by the seminal work on argumentative problem solving among stakeholders by Kunz and Rittel (1970) which, with the exception of Goel and Pirolli (1992), slipped the attention of the psychological community for many years.

The social science problems that we deal with here, are neither scientific problems nor design problems. They often involve the development of programs and projects to change the environment, people or organizations. From the area of evaluation research we used a number of important concepts to describe the argumentation underlying solutions to social science problems. The various backgrounds from which we have drawn elements to use in our research are described in chapter 2. We relate problem solving processes to demands on the communication environment and in particular on the representational notation that should support the problem solvers. This is where the main questions for our work are developed: What representational functions are needed in a computer environment that supports collaborative solving of social science problems? What is the content of a representational notation that is aimed to support problem solvers? In chapter 2 we review several different notations that have been embedded in computer-based environments. We relate their characteristics to the support that these environments offer to specific problem solving activities.

Fourth, the intended audience or user needs to be made clear; in other words who are the problem solvers that we intend to support? Is this the whole gamut of students of the Open University of the Netherlands (OUNL), as sketched above, with their different backgrounds and levels of expertise? Or is it more likely that the support offered presupposes a certain level of domain knowledge before the support can be profitably used? The environments for graphical argumentation that we will discuss in the next chapter are developed for different audiences. CSCL environments often aim to support young learners. The tools, including the interfaces and the representations they offer are developed so as to allow learns to work with these tools. The representations offered match the limited domain knowledge of the learner. In contrast, environments for professional use may offer rich and complex functionalities and representations, because they are often used in a context in which a mediator handles the technicalities so as to ensure that the professionals can concentrate on content. In the OUNL context we have to deal with novices as well as with beginning experts. The learning arrangements offered to them may use face-to-face or synchronous meetings or may be restricted to asynchronous modes of collaboration.

Fifth, and finally, we need to be more specific about the computer-based environment that we envisage. Although we started, as stated in the first sentence, with the idea of supporting asynchronous environments, this idea was quickly dropped. In chapter 2 we try to relate problem solving activities to support services and there we indicate that *some* of the activities are best performed in synchronous environments, whereas others can profit from having asynchronous modes of operation. Although the ultimate goal of the research reported here is the design and implementation of a computer-based system that supports collaborative solving of a particular type of ill-structured problems using external representations, the goals we set out here are more modest. Kirschner (2002) describes six stages in the development of a CSCL environment. The first three stages are concerned with (1) determining what the learners actually do (2) determine what can be done to support the learners and (3) determine the constraints of the learner, the learning situation and learning environment and the conventions that already exists. In the work reported here we will mainly be concerned with stages (1) and (2), leading to a set of functional requirements that will be presented in chapter 6.

Overview of this thesis

As is amply clear for the experienced reader, chapter 2 presents the major theoretical and practical underpinnings of this thesis. It discusses the psychological approach to solving ill-structured problems as well as the work on wicked problem solving. It goes on to show how the two can be brought together in a multiple agent model. A model for solving ill-structured problems is put forward that links a number of problem solving states with specific communicative and cognitive demands. These demands are subsequently translated into a number of facilities or 'services' offered by environments for Computer-Supported Argumentation Visualization (CSAV). This chapter ends with a first definition of a representational notation to support solving of social science problems.

Our first empirical study aimed to use an existing CSCL environment and make some minor changes so as to have it support solving social science problems. We started with an experiment in which we attempted to determine whether manipulating the number of objects and their semantics available in Belvédère®, a well-known environment for collaborative scientific inquiry using evidence diagrams, would trigger second year Education students to give more attention to the 'how' questions of solving an ill-structured social science problem. The experiment and its results are reported in chapter 3. Thanks (!) to a series of technical problems we learned that for these students working together using a simple representational notation was more profitable than if they had to use a richer representation. This effect was especially manifest when a richer representation was combined with a mode in which students worked on physically separate machines.

The outcomes of this experiment gave little hope that novices could work with the complexity of the representational notation that we had derived from theoretical perspectives. We therefore decided to not have a series of similar experiments, but rather concentrate on explorative studies using small samples of students from different educational backgrounds that can be considered to be beginning experts. Explorative studies were conducted using beginning experts from different domains. These explorative studies are reported in chapters 4 and 5.

In chapter 4 we describe a study where we take a first step towards the validation of the representational notation. We analyze the dialogues of teams from different backgrounds that try to solve an ill-structured social science problem and check to what extent they use concepts introduced in the representational notation. The analysis showed that many of the concepts were indeed used. It also showed some remarkable difference in problem-solving approaches between the groups which in some cases seemed to have worked counterproductively.

This led to an exploratory study reported in chapter 5. This study replicates the previous one, but changes two important factors. First, the participants, who came from different educational and national backgrounds were grouped such that they formed multidisciplinary groups. Second, and most important, these were not only beginning experts in their respective disciplines, these were also students trained in the essentials of Logical Framework Analysis. Although the scale of the study is modest, its participants are as close to the intended audience as we can get. The study showed that with these participants the representational notation can account for more of the dialogues and that more of the dialogues were on task. Further improvement seems possible however.

In chapter 6 we review the work reported in the previous chapters and then draw two different kinds of conclusions. The first part of the conclusions is in the form of global functional requirements for a computer-based system that can be used by (beginning) experts to assist in solving social science problems. The second part describes directions for further research and development.

2 Visualizing argumentation for solving ill-structured problems[1]

In this chapter we review the functionalities of a number of computer-based systems that offer visualization of argumentation to support collaborative problem solving. We will refer to these systems as offering Computer Supported Argumentation Visualization (CSAV), or CSAV environments. To review the support offered by CSAV environments we first identify a number of cognitive and communicative demands of collaborative problem solving. We discuss two traditions of problem solving and the demands that can be derived from these approaches.

The typical way in which CSAV environments aim to support problem solving is obviously in the representational facilities that they offer, but they offer communicative facilities and sometimes more advanced functions as well. To better understand the representational facilities we present and discuss a number of characteristics of external representations.

As stated in the previous chapter our concern is with social science problems that can only be solved using argumentation and informal reasoning. We concentrate on the visualization of *qualitative* argumentation and do not address the representation of formal models or simulations of the reality that the problem solvers may use. Neither will we discuss several hypertext or CMC applications. We refer the reader to a recent overview of the roots of CSAV for a more complete coverage of that subject (Buckingham Shum, 2003).

Two approaches to problem solving

Two approaches to solving social science problems are relevant to our work. The first approach is rooted in the psychological tradition (Voss, Greene, et al., 1983) while the second has its origins in work in planning and design (Kunz & Rittel, 1970).

The psychological tradition: solving ill-structured problems

CSAV is primarily used for the solution of ill-structured problems (Newell & Simon, 1972; Reitman, 1965) as opposed to well-structured problems. *Well-structured problems* (Newell & Simon, 1972) have (1) complete and unambiguous problem specifications; (2) clear criteria and procedures to evaluate whether a solution has been reached; (3) all of the knowledge necessary to solve the problem represented in one or more problem spaces with at least one problem space that can represent the initial state, the intermediate states and the goal state; and (4) an associated set of operators that can change a problem state into another state. Examples of these types of problems are the Tower of Hanoi, tic-tac-toe, theorem proving, and traditional school book problems. The problem solving process for such well-structured problems often follows a progression through three distinct stages, namely *orientation* where a problem representation is constructed, *solution*, where operators are applied to transform the problem state into a goal state, and *evaluation* where the solution and use of operators are evaluated.

[1] This chapter is based on Van Bruggen, Boshuizen, and Kirschner (2003) and Van Bruggen and Kirschner (in press).

Ill-structured problems, in contrast, (1) have an ambiguous and incomplete problem specification; (2) lack clear-cut criteria to evaluate whether a solution has been reached implying that there are no stopping rules; (3) make use of several potential information sources that may be used to represent problem spaces although it is unclear which ones should be used and how they should be integrated; and (4) have neither a complete enumeration of applicable operators nor a predetermined path from initial state to goal state. Examples of ill-structured problems are music composition, design tasks, planning tasks and management problems. When solving ill-structured problems, problem solvers do not progress in a linear way through the stages described, but rather work on partial solutions, return to refine their problem representation, evaluate, revise criteria, et cetera.

In a series of investigations, Voss and his colleagues (Voss, Greene, et al., 1983; Voss, Blais, Means, & Greene, 1986; Voss, 1991) applied a combination of Newell and Simon's (1972) information processing model and informal reasoning to problems in such diverse fields as social sciences, medicine, mathematics and foreign policy. Voss, Green, et al. analyzed the protocols of experts trying to find a way to increase crop production in the Soviet Union[2]. The problem solving process was decomposed into a *problem representation stage* where the subjects formulated the nature of the problem and a *problem solution phase*, where they solved the problem. In their analyses they distinguished between a *problem solving structure* with associated problem solving operators and a *reasoning structure* with a set of (informal) reasoning operators. Examples of problem solving operators are 'state constraint', 'state subproblem', 'state solution', 'evaluate'. Examples of reasoning operators are verbal actions: 'state argument', 'state assertion', 'state fact', 'state reason', 'state outcome', 'state conclusion' or 'state qualifier' - and operators such as 'compare/contrast', and 'elaborate/clarify'. Note, that the reasoning operators are domain specific. The analysis of the reasoning of a person who studied a juridical case used operators like 'state claim for defense', 'state fact favoring defendant', et cetera.

To solve an *ill-structured problem*, problem solvers still have to represent the problem, define constraints, apply operators and evaluate proposed solutions, but since none of these are given, they first have to draw upon other knowledge sources to determine relevant or beneficial representations, constraints, operators and criteria (problem structuring) and then justify the decisions made both to themselves and to others. The problem solver, thus, also depends on informal reasoning to establish constraints and criteria and to underpin the likelihood of success of proposed solutions.

The kinds of problems are such that the problem solver, in general, cannot be sure that the solution (s)he is formulating will actually work. It may take years before the effects of an intervention will materialize. Voss, Tylor, et al. (1983) refer to this as delayed evaluation. In other words, a proposed solution cannot, in general, be evaluated on the basis of demonstrated effectiveness, but has to be evaluated using an (informal) argumentation that will establish the likelihood of the success of the intervention as well as its acceptability. An argument cannot be evaluated in terms of whether it is right or wrong. It requires the

[2] For those readers who are younger than the author, the Soviet Union is the former northern Eurasian empire (1917/22–1991) stretching from the Baltic and Black seas to the Pacific Ocean and, in its final years, consisting of 15 Soviet Socialist Republics such as Armenia, Belarus, Estonia, Georgia, Kazakhstan, Latvia, Lithuania, Russia, Ukraine, and Uzbekistan. It had a centralized plan economy (5-year plans) that was known for its non-ability to meet the goals of the plans, especially with respect to agriculture.

evaluator to make use of other criteria such as perceived plausibility (acceptability) of the claim, perceived support that a reason leads to a claim, and the quality of the argumentation judged by taking counter-arguments into account. This informal reasoning requires the application of an argumentation structure consisting, minimally, of a claim with support (e.g., evidential reasoning). Here, as well as in the statement of the problem constraints, individuals may differ in the evaluation of relevancy or importance (Voss, Wiley, & Sandak, 1999).

The last point raises questions on how multiple agents can collaborate to solve a problem. We will return to that issue after discussing the approach to problem solving defined in the planning and design tradition.

The planning and design tradition: solving wicked problems

Based on analyses of (urban) planning and design methods and practices, and independent of the psychological research of those days, Horst Rittel - a mathematician and statistician who worked on developing socioeconomic prediction models and evaluating sociological field research - emphasized that the 'wicked' nature of problems in these domains requires an argumentative approach to problem solving (Kunz & Rittel, 1970). The kind of problems that "planners deal with - societal problems - are inherently different from the problems that scientists and perhaps some classes of engineers deal with. Planning problems are inherently wicked" (Rittel & Webber, 1984, p. 135-136). Rittel and Webber formulated a number of characteristics of such wicked problems, namely that they have no definitive formulation in that the information needed to understand the problem depends on one's idea for solving it; that possible solutions cannot be tested and revised (e.g., obviously one cannot 'try out' a trajectory for a freeway before building it); and that they have no stopping rule in that "the planner terminates work on a wicked problem, not for reasons inherent in the 'logic' of the problem. He stops for considerations that are external to the problem: he runs out of time, or money, or patience". "Solutions to wicked problems are not true-or-false, but good or bad." (op. cit., pp. 136-139). Kunz and Rittel saw the solution process as inherently argumentative, in which the problem solvers continually raise questions and argue with themselves and others over the advantages and disadvantages of alternative positions taken with respect to these questions.

Consider the following description of a wicked problem within the car industry, namely determining the features of a new car (taken from Conklin and Weil, 1997):

> 1. The actual problem is not understood until a solution has been developed. In the design of a car, some features interact with others. Adding structural support in the doors, for example, makes the car safer from side impact, but the added weight increases the cost, changes the fuel economy and ride, and requires adjustment to suspension and braking systems. Making the car safer also impacts marketing, raising issues such as pricing and demand - "How much do people really care about side-impact survivability?" And all these problems interact.
>
> 2. There are different stakeholders in how the problem is resolved. There are two clearly defined and opposing camps: the people who know what is needed (Marketing or Sales) and the people who know what can be done (Engineering or Manufacturing). Virtually all product features and design problems fall squarely into both camps. One side argues that there is no point building the product if it doesn't have Feature X; the other argues that Feature X is so expensive, complex, time consuming, untested, or otherwise impossible that it should not be tackled on this project. Management has its own stake in

> these decisions, as do many others in the organization. Some key stakeholders, such as customers and regulatory bodies, are generally not even represented in the design meetings.
>
> 3. The constraints change over time. Almost all solutions have the constraints of time (the problem must be solved before some critical date, condition, or event) and money (the solution must be cost effective). Quality is usually another key constraint. In the case of car design, some decisions, such as the addition of side-impact reinforcements, might be forced by unpredictable constraints, such as the need to impress a politician or a Wall Street analyst with the company's commitment to safety.
>
> 4. The problem-solving process ends when resources run out. Whatever is finally decided, it will be hard to claim that it was the right answer. No amount of study, laboratory experiments, or market surveys will indicate the ideal solution. At some point, the design team will have to make a decision. Inevitably, once the car is produced, critics will point out that the doors are heavy and difficult to open, while people injured in side-impact accidents will file law suits against the company.

As can be seen in the example:
(a) Wicked problems are composed of an interlocking set of issues and constraints, rather than a definitive statement of the problem itself. These problems are often not fully understood until a solution has been developed.
(b) There are many stakeholders who have expertise in different aspects of the problem to be solved, making effective problem solving more a social process than a cognitive one. Obtaining the right answer is less important than having the stakeholders accept the solution that emerges.
(c) The constraints on the solution change over time because everything one does is done at a faster pace in a changing world. Stakeholders come and go, communication is often incomplete, and rules invariably change.

Wicked problems are formally a subset of ill-structured problems: their goals are unclear (the problem is ill-defined), their search space is not well defined (and ill-structured) and neither the applicable operators nor the constraints are given. There are also a number of unique features to wicked problems however. In the first place, as stated earlier possible solutions cannot be tested and revised. Second, not even the description of a wicked problem is without implicit, often political, assumptions. Is the trajectory of the freeway a transport infrastructure problem, an economic problem, or an ecological problem? According to Rittel, wicked problems are inherently multi-disciplinary, involving stakeholders from different backgrounds, and there is no monopoly on expertise. He used the provocative term *symmetry of ignorance* to describe this state of affairs (Rittel, 1984).

Solving wicked problems thus involves different actors (stakeholders) that entertain different problem representations, and none of the stakeholders dominates the problem definition or structure. To deal with these issues, Kunz and Rittel (1970) proposed a methodology called Issue Based Information Systems (IBIS) that tried to ensure that all stakeholders could put forward their issues. Essential to the IBIS method is the absence of premature problem structuring: the argumentation structure consists of a number of topics (issues) and positions that the stakeholders hold with respect to these issues and there are no means in the method to, for instance, create problem decompositions. IBIS contains none of the domain dependent operators that Voss c.s. used to describe the problem solving behavior of experts and novices. Several CSAV environments such as gIBIS (Conklin & Begeman, 1988) Mifflin® and Compendium® (Selvin, 2003) were inspired by Rittel's IBIS method.

Rittel also added an important aspect to the problem context, namely that there are many stakeholders whose views on the problem may vary (Rittel & Webber, 1984). Thus, the idea of multiple agents with multiple representations was implicit in this seminal perspective. In psychological terms, Rittel described a problem solving context where multiple actors, having different representations of the problem, are trying to solve the problem. Understanding the role of CSAV environments in the problem solving contexts that Rittel sketches forces us to rethink how informal reasoning and argumentation proceed if more agents from different backgrounds enter the scene. Somehow, their different views, knowledge and operators have to be coordinated. This brings us to the question how multiple agents, using multiple representations can collaborate to solve problems and how visualizing argumentation fits into this problem solving process.

Coordination problems for multiple agents

There are several reasons to assume that collaborative problem solving will only work if the problem solvers succeed in coordinating their problem representations, their data, and the operators that they use. Visualization of argumentation can facilitate this problem solving in a number of ways. First, it is a means of explicating and sharing representations among the actors which may help them to build the partially shared representations that may be essential for collaborative problem solving (Alpay et al., 1998). Second, it may help the problem solvers maintain focus, one of the success factors of collaborative problem solving. Finally, it may help problem solvers maintain consistency, accuracy and plausibility – three important aspects on which solutions are evaluated (Alpay et al., 1998).

The amount and kind of knowledge and skills that must be coordinated between agents are the sources of problems that make cooperation problematic. Superficially, it may seem that involving more people in problem solving increases the chance that someone has the knowledge to solve the problem. Unfortunately, incomplete understanding and misunderstanding between agents are the more likely outcome, especially if the agents have varying levels of expertise within a single domain or if they come from different backgrounds. In these cases the agents will have different *problem representations* manifesting themselves in different data, formats and operators that they apply. Agents may also differ in the *criteria* they apply to evaluate solutions and the arguments that underpin claims.

Misunderstandings due to differences in representation are not easy to avoid, especially between experts who have spent thousands of hours mastering their fields and novices whom they may be working with. Discussions between experts and novices have been shown to be notoriously difficult (Bromme, Nückles, & Rambow, 1999). Specific issues cannot be explained at the required level of understanding to a novice, and when the expert tries to explain more, the novice is overwhelmed. This is not only due to the nature of the knowledge to be conveyed, but also because experts tend not to be able to 'tune in' to the level of novice understanding. This is not only true when dealing with very different fields like law and microbiology where one person can be an expert in one field and know absolutely nothing about another. Boshuizen and Tabachneck-Schijf (1998) found that cooperation problems also occur when different experts have fields of expertise that partially overlap and partially differ in content, concerns and/or paradigm. They call this 'distributed or multiple representations in multiple agents' and define it as the circumstance where multiple human or artificial agents have dissimilar representations about an object, person, interaction or situation. They describe how representations may vary along the dimensions of *data*, *format* and *operators*.

When people think differently about an object or have different perspectives on it, their representations vary along the *data* dimension. Data is what we call content in everyday life. Since the content mastered by an expert is vast and highly integrated, it is virtually impossible for a novice to learn that which is relevant for solving a certain problem. This is compounded by the fact that even when two people have mastered the same concepts, they still might not understand each other. Different people can have different perceptions or have different prototypes of the same concept. This is painfully apparent in the differences exhibited between the understanding of concepts via 'real physics' and 'lay physics', a phenomenon often labeled *misconception*. Other differences stem from different domain-specific representations: stakeholders of a wicked problem often represent problems in different ways, thus a problem in the production of a lawn mower is a logistics problem for the planner, a design problem for the industrial designer, a personnel problem for the human resources manager and a financial problem for the accountant. Recognizing such differences at the conceptual level can often take more time, result in more surprises and be more of a nuisance than is the case when people entirely lack each other's concepts. Different contents of representations can also be the source of many misunderstandings, for example the scheduling of a meeting between an English and a Dutch delegation at 'half ten', which in Dutch means 9:30 and in English 10:30.

A second dimension along which representations may vary is *format*, for example when one representation is propositional and the other visual (e.g., mathematical notation vs. Venn diagrams). Note that argument visualization is a means of forcing the use of the same format, or surface representation as Stenning (1998) calls it, on different people. Stenning points out that using the same surface representation may help to unravel differences in conceptual structures.

The third dimension reflects that representations may differ regarding the *operators* applied. Differences along this dimension may appear unexpectedly and may lead to conflicts because one person thinks that the other's reasons are 'unfair'. Boshuizen and Tabachneck-Schijf (1998) give an example of the use of different sets of operators: a legal approach toward solving a misunderstanding between people versus using more common-sense ideas about solving it.

Another source of misunderstanding stems from differences in the *criteria* that agents use to evaluate solutions and arguments. While a blueprint for a containment construction may convince engineers that the container is an acceptable solution to an environmental problem, it is not likely that an environmentalist will accept this as a solution. Differences in criteria between agents can be so fundamental that they lead to the complete rejection of specific approaches to problem solving or to certain reasoning patterns. An example is an abortion rights activist who considers life to begin at birth and a member of the clergy who considers life to begin at conception, if not earlier.

Data, formats, and operators can form extended procedures or lines of reasoning (macro-operators) which are needed for problem solving. We define such macro-operators as sequences of operators learned during past (problem solving) experience which can be shared by domain area experts (i.e., mathematical procedures, juridical reasoning). Argumentation strategies for a specific line of reasoning reflect common values and goals for solving problems in a specific domain and the role of argumentation therein. Scientific researchers, for example, attempt to *achieve consensus* based upon the exchange of arguments that are open for critical debate while lawyers try to *win a conflict* by convincing a ruling authority that their claims should be honored above those of the other party. However, even within a discipline, striking differences can be found in argumentation. More formal reasoning styles, for example, are tightly linked to scientific paradigms and are often not accepted by scientists with a different background.

What this suggests, is that there are serious coordination problems when dealing with multiple agent, multiple representation situations. In the next section we try to formulate the coordination needs as part of the broader cognitive and communicative demands of collaborative problem solving. Before detailing these demands, we define a number of important preconditions and make a distinction between three stages of problem solving through which problem solvers move back and forth.

Cognitive and communicative demands of collaborative problem solving

Preconditions

Before specifying cognitive and communicative demands put on CSAV environments we spell out a number of more or less social preconditions underlying collaborative problem solving, namely certain minimal levels of *shared understanding*, *accountability* and *trust* that must exist before any collaborative problem solving may occur.

The first precondition is a minimum of *shared understanding*: Shared understanding is the state where two or more people have equivalent expectations about a situation: Their explanations of the situation and their predictions for how it might develop are the same. In the lawn mower example, the agents share the understanding that there is a problem which they have to solve together and which they can solve together, namely that a lawn mower needs to be manufactured. Before they can tackle the problem of how best to manufacture (and market) it, they must first develop a minimum shared understanding of how the problem can be represented and which operators and reasoning schemes are admissible for solving the problem (as in the example of the linguistic vs. psychological explanation).

A second precondition is *accountability.* Accountability is the social mechanism underlying responsible behavior between people, for example that one team member does not plagiarize a fellow team member, take the credit for work done by another team member or work to the disadvantage of a fellow team member. This precondition does not imply that agents necessarily have the same goals. Even in situations of adversarial collaboration (Cohen, Cash, & Muller, 2000) parties adhere to standards for exchanging and sharing views and information.

A final, but also very important factor is *trust*. Trust is the perceived ability to rely on the character, ability, strength, or truth of someone or something and is the deciding factor in a social process that results in a decision by an individual to accept or reject a risk based on the expectation that another party will meet the performance requirements (Zolin, Fruchter, & Levitt, 2002).

Problem solving states

Research on problem solving distinguishes three general states in problem solving (Newell & Simon, 1972) namely an *orientation state* to determine what the problem is, a *problem solving* *state* which entails the actual problem solving process, and an *evaluation state* which determines both whether the problem has been solved and whether the 'right' problem has been solved. Note that each state needs information that is often lacking in ill-structured or wicked problems. In such problems, the problem description and constraints are ambiguous at best; the applicable operators are not given, and the criteria to evaluate solutions are absent. This straight-forward, linear process through the three stages is only possible when solving well-structured problems using well-known algorithms. When dealing with ill-structured problems, problem solvers move back and forth between orientation, problem solving and evaluation. Trying out a partial solution (or rather a mental model of it), for example, may help the problem solvers to better understand the problem and may lead to adding or refining constraints and criteria (Norman, 1998). Visser (1990) described the activities of software designers as being guided by a data-driven and opportunistic control strategy rather than as following a plan or routine as reported by the problem solvers themselves. Although Goel and Pirolli (1992) found a somewhat more linear progress through sub-states of the problem solving state, they describe the control strategy as 'limited-commitment-mode control' which allows the designer to put work on an unfinished design module on hold at any time and redirect attention to other modules or tasks.

In the *orientation state*, participants structure the problem, clarifying the problem description, constraints and criteria for solution and evaluation. Participants ask themselves whether the actual problem is the one stated (surface level) or whether the real problem lies deeper, hiding behind the problem as it appears to them. Participants also need to clarify, and often state themselves, the boundary conditions and constraints, the context(s) in which the problem is embedded and who the owners of and stakeholders in the problem are, and so forth. Using the results of intermediate and partial solutions, the problem solvers will review and revise their understanding of the problem, its constraints and the criteria to evaluate a solution. In the *solution state* participants plan how to solve the problem and then try to execute the solution in a more or less systematic way. As previously described, this is often a data-driven process with a loose control strategy that lets problem solvers move back and forth between partial solutions, problem structuring and evaluation. In the *evaluation state* both process and outcome will be checked, that is the participants not only check the correct application of the operators, but they also determine whether the problem as intended has been solved. Since there is no clear-cut criterion to evaluate solutions or the correct application of operators in ill-structured problems, the problem solvers have to define, review and revise their criteria.

Orientation seems crucial for collaborative problem solving, because the representations of the problems which are made here are closely linked to the type of acceptable argumentative reasoning that can be used in the solution. In multidisciplinary teams, differences of interpretation of what the problem is and of what the problem solving strategy should be may lead to great difficulties if the different viewpoints do not converge. This is the point – in our opinion - where argumentation is most important and where the design and use of CSAV plays the greatest role. Cognitive and communicative demands vary across the states of problem solving (Duffy, Dueber, & Hawley, 1998) and these variations have important consequences for the design of CSAV environments. In their first efforts to structure a problem, the problem solvers often have different or only partially overlapping representations of the problem and need to work on at least a minimal shared understanding of the problem. The communication between the problem solvers is relatively unfocused and issue-based, leaving room for them to explore interpretations of the problem, the constraints, and so forth (Duffy et al., 1998).

Others question whether the use of argument visualization tools such as external representations in this stage will facilitate the problem solving process. Van Gelder (2003) avoids using argument visualization in the early stage of problem representation while Selvin (2003) starts using visualizations immediately, which has the additional advantage of creating a record of the problem structuring activities. Whatever the preferences, if external representations are used at this stage, then their representation of the argumentation has to be non-committal, that is the representational notation must allow expression of different perspectives on the problem.

One might also question whether the notations used to express the solutions should also be non-committal. Voss c.s. seem to hold that since problem solvers use domain-dependent operators and macro-operators when solving problems, the solution process carried out should not be hampered by representational notations that lack such expressive power; a view that encourages the use of domain-specific representational notations and discourages the design and use of generic tools. But there is a trade-off here. In a situation where several agents are involved and where we want to use external representations as a means of mediating the problem solving process, the representation has to be such that it can be, in Selvin's words, 'put in the middle'. This requires that the representation be sufficiently understandable to all actors, and thus not *too* domain-dependent.

A direction towards a resolution of this apparent contradiction can be found in Alpay et al. (1998) who - in studying how interdisciplinary teams of engineers and psychologists used multiple representations to analyze traffic accidents - identified a number of characteristics and dimensions of these representations. One dimension is a *permanent – temporary* axis, where permanent representations correspond to systems, procedures and models that the experts use on a *regular* basis in *standard* situations and temporary representations are representations built dynamically during the analysis of a *specific* accident. A second dimension is a *shared – unshared* axis where, for example, a simple functional model of the driver involved in an accident was shared by engineers and psychologists, while a richer version was only shared between the psychologists in the team. These two dimensions are extremely important, since permanent and shared representations seem to be the core of the representational coordination. If we are to design and develop powerful CSAV environments, we cannot neglect that along with the permanent, shared representations, these environments should also allow for temporary, unshared representations. Another important coordination mechanism that they found relates to the distinction between *control representation* and *topic representation*. Control representations are representations that guide operations on topic representations such as models, phase decompositions, et cetera. Selvin (2003) makes similar distinctions and formulates requirements such as *preservation* (i.e., maintain information as well as conceptual frameworks), *rigor* (i.e., support rigorous methods) and *repeatability* (i.e., afford easily reproducible methods) all of which hint at common structures and control representations. Selvin's requirements not only stipulate the re-use of structures such as tasks, Toulmin structures, or stages, but also the definition and representation of emerging structures. He provides examples of support built into the tools such as using databases to maintain consistency and the continuity. He also formulates relevant requirements. His reframing requirement can be seen as a means to (re)build consistency and plausibility. We should, however, keep in mind that his description of operators and requirements have been generated on the basis of studies in a professional context, that is with experts, and experts are known to have available more, accurate representations which they can switch between with great flexibility.

The style of communication during solution formulation is more *focused* and *topic*-based (Duffy et al., 1998; Goel & Pirolli, 1992; Van Gelder, 2003). The capability to maintain focus is one of the major success factors in collaborative problem solving (Barron, 2000; Erkens, 1997; Veerman, 2000). Maintaining focus and coherence is often problematic, especially in asynchronous discussions where discussions tend to get scattered and lose coherence (Herring, 1999).

Evaluating whether the problem as intended has been solved, whether the operators that have been used have also been applied correctly and whether the constraints have been met is a continuous process when solving ill-structured problems. This process will not only lead the participants back to the solution states, but will also lead to reconsidering the evaluation criteria as well. A strategy often applied in this oscillation between solution and evaluation is *satisficing* (introduced by Herbert A. Simon in Models of Man, 1957: To obtain an outcome that is good enough; as opposed to maximizing which seeks the biggest, or optimizing which seeks the best) where the problem solvers repeatedly evaluate their (partial) solutions against the constraints, leading to continual refinement. Solutions often seem to emerge more or less effortlessly out of this process (Voss et al., 1999). An important set of criteria to evaluate the argumentative underpinnings of proposed solutions are those characteristics which are used to achieve representation-management goals, namely, maintaining consistency, accuracy and plausibility of the representations (Alpay et al., 1998). These criteria are quite consistent with those formulated by Voss c.s. Before we consider how CSAV environments might help to meet the cognitive and communicative demands of the three stages of the problem solving process we recap them in Table 2.1. In the following sections we will show how the cognitive and communicative demands of collaborative problem solving may be translated into requirements on CSAV environments. We focus on representational features of CSAV environments.

Table 2.1
Cognitive and Communicative Demands of the Problem Solving States

Problem solving states	Cognitive demands	Communicative demands
Orientation	Problem Representation Constraints Problem structuring Establish shared representations	Issue-based communication Brainstorm Build trust Establish common ground
Solving	Apply macro-operators to produce solutions Use topic and control representations Maintain coherence Maintain accuracy Maintain plausibility	Topic-based discussion Maintain common ground Maintain focus Conflict detection and resolution Knowledge negotiation
Evaluation	Evaluate solutions Evaluate constraints Evaluate process	Negotiate criteria

Characteristics of external representations

CSAV environments aim to support collaborative problem solving by making shared external representations of their argumentation available to the problem solvers. These external representations are construed using a limited set of objects and relations and adhering to certain rules on their use and combination. To this end, Suthers (2001), distinguishes between *representational notations* for representing the objects and relations, *representational tools* for implementing the notation, and *representational artifacts* which are the products constructed with the tools.

The *representational notation* contains the primitives - the objects and relations that can be used in the representation of the argumentation and the rules that govern their use. Such a notation may define objects like claim, data, warrant, as well as the type of relation between these objects and associated features like strength of belief, hierarchy, or causality. Finally, the notation specifies which combinations of objects and relations are allowed. It can stipulate, for example, that data can be related to one or more hypotheses and that no relation is allowed between hypotheses. Before such a representational notation can be used it must be (and is) implemented in a piece of software, the *representational tool*. Several design decisions are then made to implement the notation such as the choice of symbols used to denote the objects and the relations and tools used to handle them. A decision, for example, needs to be made with respect to what would happen if the user (following the example used before) tried to relate two hypotheses? Would the system allow this, ignore it, or ask/force the user to remove the relation? Functionalities can be added to the tool (e.g., focusing and zooming) and future usage considerations are taken into account (e.g., if the tool will be used in a synchronous or asynchronous setting, whether users will work alone or make use of joint workspaces). Using the representational notation embedded in the representational tools, the users create the *representational artifacts*: the argument maps, Toulmin structures, evidential diagrams, et cetera.

A number of characteristics of external representations were identified by De Jong et al. (1998). Their work aims to cover a broader gamut of external representations than we need and we therefore use only a subset of their characteristics.

Ontology

The *ontology* of a representation defines what can be seen in the domain represented and how it will be seen. The ontology "refers to the *content*, to the objects and relations one uses to represent a domain, not so much to the symbols by which objects and relations are denoted" (De Jong et al., 1998, p.11). Selvin (2003) refers to this characteristic as the 'depth of palette' and indicates that having a smaller set of objects and relations can be an advantage. The development of the Belvédère® environment (Suthers, Weiner, Connelly, & Paolucci, 1995; Suthers, 1999) corroborates these observations. Early versions of the Belvédère® environment were developed as a tool to represent the argumentation in scientific inquiry and had a rich set of objects and relations. Learners, however, were disturbed by the richness of the representation and were often lured into off-task discussion. Later versions of Belvédère® have a smaller set of those objects and relations so that learners may better concentrate on the core concepts. Alpay et al. (1998) as well as Selvin (2003) indicate that the ontology of CSAV environments needs more than primitive objects and relations. It also requires complex structures such as the components of a task or a Toulmin argument structure that can be used as frames to represent interrelated data or as process control structures. These authors also stress the importance of re-use of these structures, thus in Selvin's terms enabling rigor and repeatability.

Perspective

Closely related to the ontology is the *perspective* or view. This term is multivalent, and can be related to views on systems from, for instance, a functional, a behavioral or a physical perspective (see De Jong et al, 1998 for examples). In CSAV the term is most often used in the sense used by Stahl (2001) for describing the different conceptualizations of a problem. One stakeholder, for example, may have an environmental perspective on a problem, whereas another may conceptualize the nature of the problem as technological or economic. Differences between these perspectives correspond to differences between conceptual representations (Stenning, 1998), that is the underlying conceptual systems are different.

Specificity

A third characteristic is *specificity*, defined as "the demand of a system of representation that information in some class be specified in any interpretable representation" (Stenning & Oberlander, 1995 p. 98). The specificity of a representation may require disambiguation: we cannot, in a diagram, represent that an object is 'next to' another object, it has to be represented at a particular locality. Thus, graphical representations limit abstraction and aid processibility of the information. We will not use the concept specificity in the formal sense used by Stenning and Oberlander, but following Suthers (2001) we will apply the concept to denote the *categorical choices* that the representational notation forces the user to make. For example, in Belvédère the specificity enforces the learner to make the basis distinction between theory and data.

Precision

Another characteristic of external representation, and one which is strongly related to the representational notation, is the *precision* or accuracy with which the representation reflects the underlying model. In science and mathematics this is related to quantitative vs. qualitative models. In other domains it will reflect the nature of the objects (e.g., differentiating between a hypothesis and a prediction) and their relations (e.g., making various relation types available). Obviously, the precision of a representation is limited by the ontology. A limited set of objects and relations does not allow detailed, precise statements in most domains. It is difficult to draw a line here for determining how specific a CSAV environment should be, although Alpay et al. (1998) point to a number of interesting directions that can help, including the use of shared (partial) models.

Modality

The last characteristic of external representations is its *modality*, the form of expression used for displaying information such as text, animation, graphs, et cetera. The modality corresponds with the way the representational notation is implemented in the representational tools. Thus notational systems with the same set of underlying concepts and relations (the ontology) may be expressed, for instance, as graphs, hypertexts or feature-comparison matrices. Suthers (1999) has drawn attention to the effects that these different notations may have on the learner discourse and ultimately on learning, and as he points out, these effects may even be greater if learners operate in a joint workspace.

CSAV environments

Services of CSAV environments

CSAV environments offer a number of services to their users. We make a distinction between *representation services* that enable the users to prepare an external representation; *communication services* that enable the users to share representations and exchange information and finally, *guidance and support services* where the system takes the initiative to point the users to issues that need further attention. The representational services correspond to what Suthers (2001) calls the representational tools: an implementation of the notation.

Suthers' Belvédère® (version 2) environment is an example of a system that provides all three types of service. It offers a number of *representation services* in that users can put together diagrams using a limited set of objects and relations and can express how certain they are about these objects and relations. It offers *communication services* such as inquiry diagrams which are created in a shared workspace so that students may share their work while at different locations and a chat window real-time communication. It also allows individual students to open a previously shared joint discussion and add to it, thus working in an asynchronous mode. Finally, Belvédère® offers *guidance and support services* such as invoking an intelligent coach that inspects the structure of the diagram and pinpoints certain weaknesses in the diagram. The coach highlights the parts with the weaknesses (i.e., uses a communicative function), presents its diagnosis, and offers advice on how to improve the diagram (see Figure 2.1).

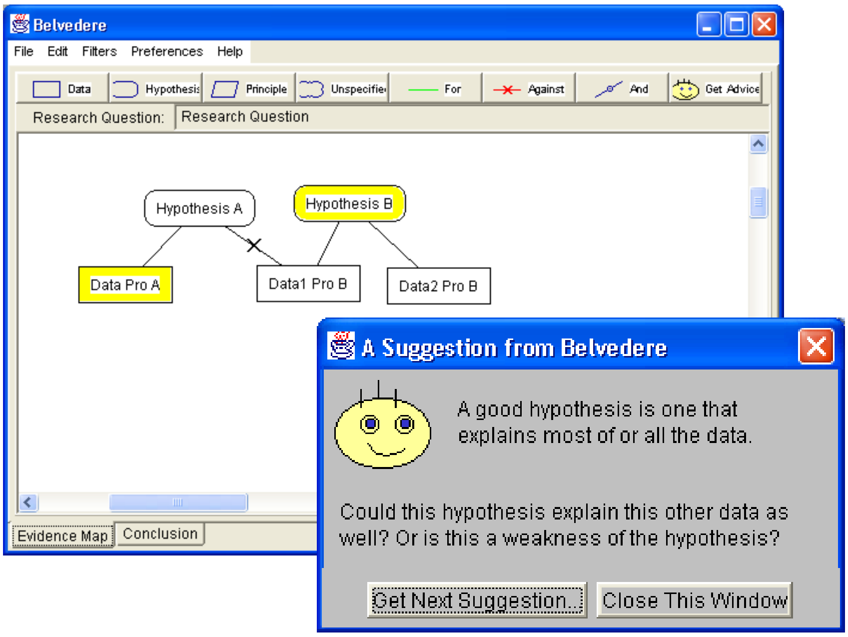

Figure 2.1. Belvédère® coach offering a diagnosis and remedy.

Translating demands to services

The cognitive and communicative demands of collaborative problem solving formulated above can be translated, although they may be hard to implement in combination, as requirements on the services offered by the CSAV environment and, as far as the representational services are concerned, as requirements on the characteristics of the representational notation. We will present these in more detail when we discuss CSAV environments. Here we present some of the more salient issues in translating the demands

Synchronous communication services allow participants to freely exchange ideas as loosely related issues during *orientation*. Here initial ideas about causes and possible solutions are generated. When different viewpoints on the problem exist, the representation services should not be biased for particular views. Representations that meet those requirements are IBIS (see Figure 2.7) and more general models to represent argumentation (but note that these as well are biased against formal models, for example). These are weak methods to represent a domain and it is likely that they trade-in accuracy for applicability. All CSAV environments assume that, eventually, a substantial overlap in the viewpoints on the problem can be realized. Maintaining a wide variety of viewpoints for a prolonged time in a computer-based environment is still extremely complex (Stahl, 2001).

For the *solving* state a first communicative demand required that the solvers can now focus on topics to explore in depth. Asynchronous communication may be best suited to meet these demands (Sloffer et al., 1999) because the activities in this state require reflection, gathering information, and deeper exploration. A second demand is to maintain task-oriented collaboration, that is maintain common ground, focus, find areas of conflict, and negotiate or co-construct knowledge. This requires frequent evaluation and coordination best done in synchronous mode since synchronicity allows swift reaction and repair to emerging problems or conflicts A combination of asynchronous communication with external representations has been found to help learners maintain focus and coherence (Herring, 1999).

Maintaining coherence, accuracy and plausibility translate to representation services. Some of the demands formulated here translate to complex requirements on representation services. Control representations (Alpay et al., 1998) guide the solvers through a series of topics. Most of the systems do not offer process support. As will become evident in the next chapters, this may be one of the crucial supports to add to a CSAV environment. Partially shared representations require at least a similar sort of administration as plausibility. Their visibility for others may become problematic. Some systems (e.g. Crocodile®) go further and offer guidance and support services that warn users that coherence is being lost or that plausibility has dropped below a certain threshold (Wessner, Pfister, & Miao, 1999).

Evaluation, finally, uses a number of criteria (such as plausibility) already discussed and examples of systems that support these criteria will be given in the next section. Other criteria, such as 'the problem as intended was solved; the correct operators were applied correctly' are not supported by the majority of CSAV environments.

Types of CSAV environments

The ultimate goal of this research is the design, development and implementation of a tool for CSAV in a CSCL environments which can be used for ill-structured and wicked problems in an educational setting. For this reason we briefly discuss some CSAV-environments developed for use in different settings, in particular for CSCL and planning and design. For a more in-depth discussion of different CSAV-environments the reader is referred to Buckingham Shum and Hammond (1994) and Van Bruggen and Kirschner (in press). The systems themselves are of varying nature (from industrial strength to research prototype) and have to serve different users and different purposes (from product design in multi-nationals to touchy problems in local government and community to collaborative learning in schools).

There are, for example, systems and methodologies like Compendium® (Selvin, 2003) that are used to support problem solving in synchronous business and public administration settings, facilitated by experts in using the system. These are settings where professionals - experts in their areas - solve problems and where a number of requirements formulated by Alpay et al. (1998) apply. These will not be discussed here.

In contrast to this synchronous, facilitated mode of operation to support problem solving by experts, argumentative CSCL systems are usually operated by students, novices or beginning experts, who *learn* how to solve problems using argumentative devices. Their activities are not facilitated but are - at most - supported by a teacher. The CSCL context puts a number of additional demands on the functions, the interface and the representations used. According to Kolodner and Guzdial (1996), the interfaces emphasize *structure* for novices who need guidance to succeed and *flexibility* for students with diverse needs, the systems offer (multiple) representations that lend themselves to extrapolation and discussion, and the software can support different logistic functions in that it can be used synchronously or asynchronously for either local or remote collaboration.

Kolodner and Guzdial (1996) further note that CSCL systems can fulfill a number of *roles*, namely: (1) promoting inquiry and sense-making; (2) facilitating knowledge building by providing a forum for collaboratively presenting arguments, raising learning issues, and reaching consensus on new knowledge; (3) keeping records and/or functioning as an external memory; (4) enabling communication with distant communities; (5) promoting reflection of alternative perspectives, solutions, and critiques; and (6) supporting teacher planning and implementation of collaborative activities. Except for the support of teacher planning, these seem to be roles that may apply to CSAV environments in general.

CSAV environments in CSCL

Bell (1997) makes a distinction between discussion-based tools and knowledge representation tools in CSCL. Discussion-based tools support dialogical argumentation in a group while knowledge representation tools not only support the dialogical argumentation, but also support the representation of the argumentation by the individuals. (Suthers & Weiner, 1995; Suthers, Toth, & Weiner, 1997)

Discussion based tools

Discussion-based tools offer students an asynchronous environment in which they can exchange arguments. The structure of the argumentation is not explicitly represented, but is usually embedded in the threads of an electronic discussion.

One of the best known discussion-based environments is CSILE® (Scardamalia, Bereiter, Maclean, Swallow, & Woodruff, 1989; Scardamalia & Bereiter, 1994), now commercially available as Knowledge Forum®. In Knowledge Forum® (see Figure 2.2) students use a communal database of interrelated text and graphical notes. When working with a recognized problem, students are required to enter notes with an identified type of content: 'My Theory', 'New Information', 'Comment' or meta-cognitive notion: 'I Need To Understand'. Notes are related by links like 'References', 'Build-On', and 'Quotes'. A special note is the 'Rise-Above' note that subsumes a number of nodes. All notes are entered in the communal database and are available to other students for search and comment. Knowledge Forum® offers a number of views (diagrams, maps) to which a network of related notes can be attached.

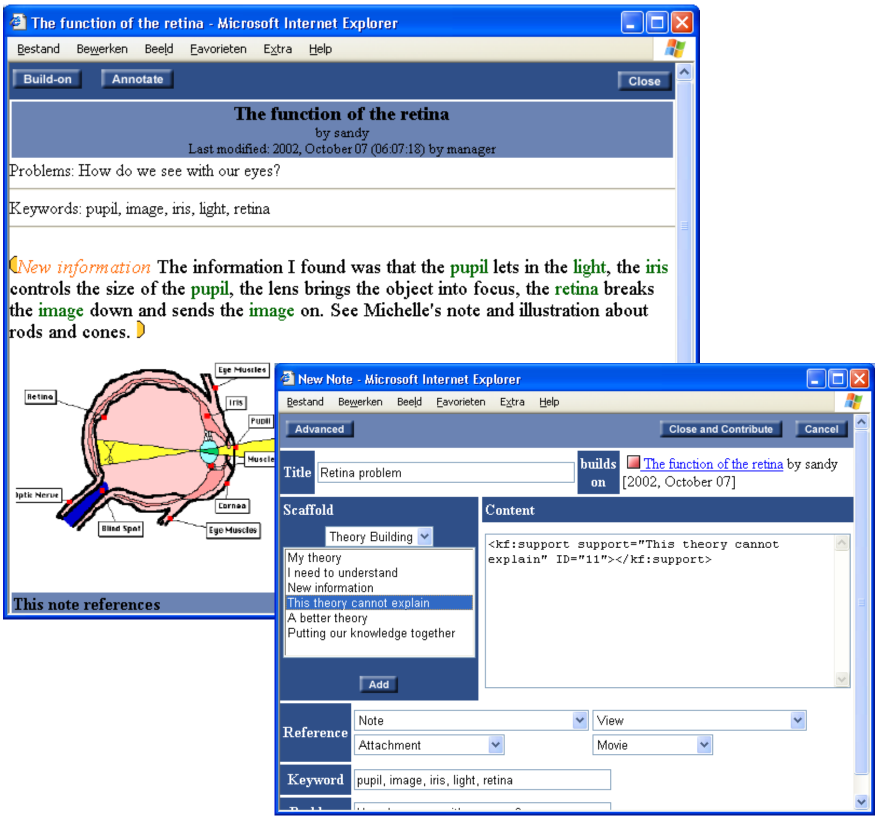

Figure 2.2 Knowledge Forum®.

Another well known environment is Collaboratory Notebook® (Edelson & O'Neill, 1994; Edelson, O'Neill, Gomez, & D'Amico, 1995) which is a shared workspace where learners can enter pages and relate them to each other through the use of hyperlinks. There are several types of pages, including questions, conjectures, evidence for, and evidence against. In the table of contents for a Collaboratory Notebook® (see Figure 2.3) the icons represent different types of nodes and the indentation represents the threads in the discussion. Discussion in the Collaboratory Notebook® is scaffolded by an interface that monitors the semantic links between nodes. Scaffolds suggest particular follow-up pages (e.g. an 'evidence for' page as a follow-up to a 'conjecture page'). In this way the structure of the argumentation is modeled to the students.

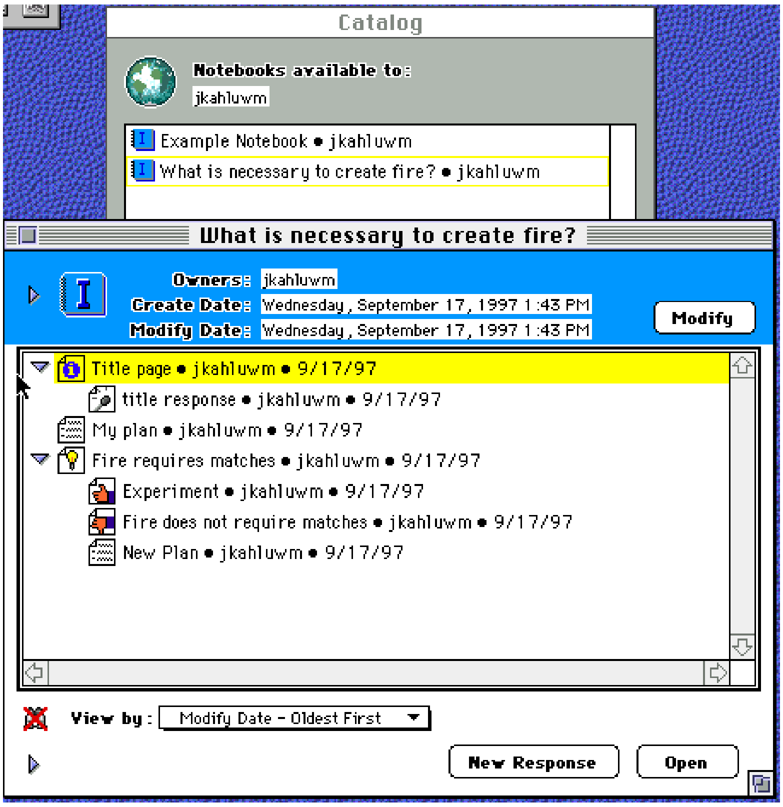

Figure 2.3. Collaboratory Notebook®.

The major characteristics of these environments that make use of discussion-based tools is that they offer functionalities to engage learners in an asynchronous discussion. By the use of particular types of nodes the nature of the contribution to the discussion can be signaled to others. The content of the contributions is contained within the nodes which limits the possibilities of more advanced services.

Knowledge representation tools

Two well-known examples of knowledge representation tools are SenseMaker® (Bell, 1997, 2001) and Belvédère® (Paolucci, Suthers, & Weiner, 1995; Suthers & Weiner, 1995; Suthers et al., 1997). SenseMaker® (Bell, 1997; Bell, 2001) is that part of the Knowledge Integration Environment / Web-based Inquiry Science Environment - KIE/WISE® (Bell, Davis, & Linn, 1995) developed for inquiry based learning in the natural sciences. KIE is a learning environment that uses the Internet to help middle and high school students develop an integrated understanding of science and a critical eye toward the complex resources found on the Web. It is the product of extensive research and classroom trials exploring innovative uses of the Internet and World Wide Web for K-12 science instruction. It supports learners in investigating rival hypotheses by guiding them to claims and evidence. SenseMaker® (see Figure 2.4) collects a hypothesis and its supporting or refuting evidence in a so-called ‘claim frame’. Evidence that supports one explanation while refuting another is not explicitly signaled. SenseMaker® offers a basic agenda function (see the ‘to be sorted’ list in Figure 2.4).

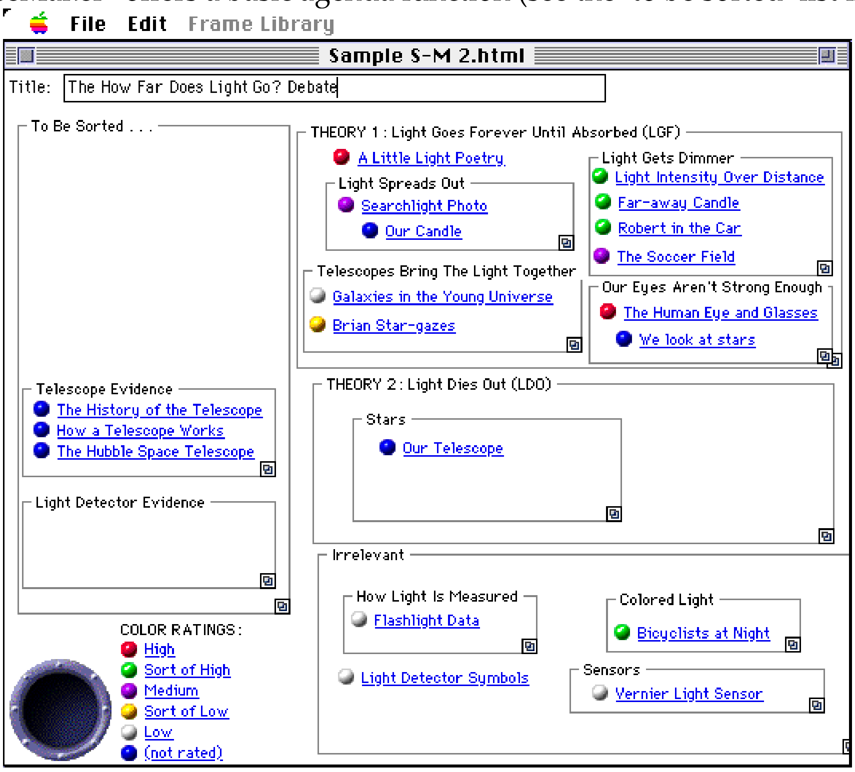

Figure 2.4. SenseMaker®.

The Belvédère® environment is also committed to scientific inquiry. Its representation and communicative services aid learners in formulating scientific explanations in ‘evidence maps’. The *ontology* of the evidence maps is implemented in the objects and relations that students may use in creating those maps (e.g., principle, hypothesis, data, and unspecified). The relations are reduced to a basic set of ‘for’, ‘against’, and ‘and’. Students can express the strength of their beliefs in the objects and relations. Belvédère® supports discussion of rival hypotheses by linking evidence (data) or by undercutting the link. It does not support direct refutation of data.

The *specificity* of the representational system of Belvédère® forces learners to indicate which type of object they add (Suthers, 1995), not allowing entry of objects of a ‘not-yet-decided’ type. This has often led to epistemological discussions between learners rather than task related discussion. Suthers notes that a weaker representational structure (e.g., one with a ‘not-yet-decided’ type) could evade the issue, but would leave room for different interpretations of the representation[3].

[3] There were a number of drawbacks associated with the earlier representational scheme of Belvédère®. The early versions had far more objects (theory, claim, warrant, observation, law) and relations (supports, causes, explains, then, AND, conflict, negates) than current versions and it offered a very explicit Toulminian perspective on argumentation. It was noted that the detailed level at which relations could be presented led to interference with the task and sidetracked students to deal with non-goal tasks (Suthers, Toth & Weiner, 1997). In subsequent versions of Belvédère® (version 2 until 4) the number of objects and especially the different types of relations (precision of the representation)

A third example is the Text Composer, Computer supported & Collaborative (TC3® - Figure 2.5) environment for collaborative writing (Kanselaar et al., 2003) which integrates a number of tools, including a diagram tool that represent the argumentative structure to be further developed in the text. The diagram contains diverse nodes (Information, Position, Argument (pro and contra), Support, Refutation and Conclusion) that correspond to a more general purpose argumentation.

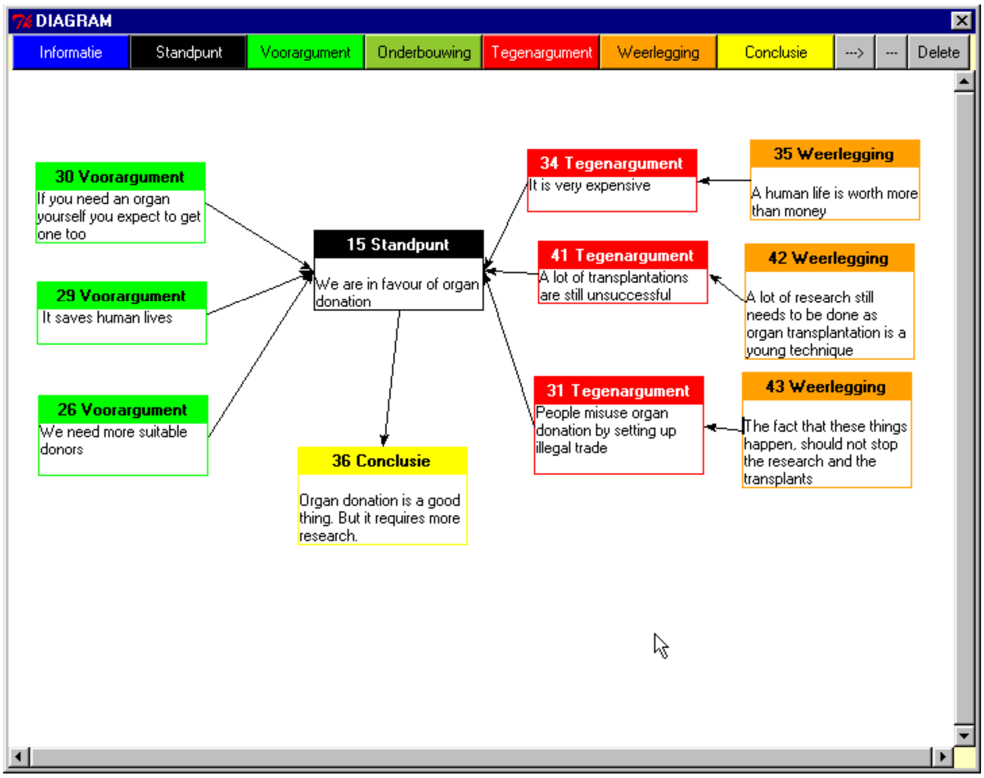

Figure 2.5. TC3® Diagrammer.

The major characteristics of these environments that make use of knowledge-representation tools is that they offer synchronous means of explicit argumentation and that the type of content is known to the system. This allows more advanced services. The Belvédère coach is an example in case: because the systems knows the structure of hypotheses and data entered into the diagram, it can give feedback and hints.

Representation services encountered

Most of the CSCL environments presented here are concerned with scientific explanations. The systems were developed for children or young students and therefore use a basic set of objects and relations to express and discuss scientific explanations. This basic set also limits what can be expressed and how accurate. There is, for example, no object 'prediction' in Belvédère®.

Knowledge representation tools offer a single representational system for expressing scientific explanations and the vocabulary offered is confined to that domain. There are no facilities for expressing (partially) shared representations or different perspectives. Despite this limitation, graphical representations make loss of coherence easy to spot, at least in a

---

were reduced by eliminating redundant relations and concentrating on those topics that were considered core to learning (make a distinction between facts (data) and suppositions (hypotheses).

superficial way. There is also some evidence that working with graphical argument representations improves *focus* (Veerman, 2000). One of the most obvious ways in which such graphical environments offer assistance in maintaining focus is by allowing deixis (i.e., the ability to point to or highlight the topic under discussion). In the Belvédère® environment the coach can help students to *focus* on those parts of the evidence diagram that need further work as described above.

Discussion-based tools focus on the *types* of entry –message or node – that can be entered and how these entries can be combined. The Collaboratory Notebook® allows students to take a number of predefined steps, but there are limits to the actions that can be taken. One cannot, for example, respond to a hypothesis entry by adding another hypothesis entry. On the other hand, discussion-based tools impose virtually no limitations on what can and can't be expressed in the entries. This freedom to enter almost any text in the messages or nodes makes conflict detection and thus knowledge negotiation very difficult. *Conflict detection* and *knowledge negotiation* are considered important in CSCL environments, because conflict is seen as an important trigger for collaborative argumentation in which learners engage in knowledge negotiation and eventually in knowledge construction. Miao, Holst, Holmer, Fleschutz and Zentel (2000), for example, point out that the information needed to spot the location of conflicts, individual comments or opinions is hidden in textual descriptions in many text based CMC-systems. This means that, in general, there is no way for the system to know where a conflict exits, let alone offer support to resolve it. Their Crocodile® system collects measures of 'degreement' - degree of agreement – to measure shared knowledge as well as knowledge where participants disagree.

Maintaining *coherence* and *focus* in discussion-based tools, or computer-mediated conferencing systems is also problematic. Discussions tend to lose coherence (Herring, 1999). Hewitt, Scardamalia, and Webb (1997) implemented the 'Knowledge Map': a tool that allows CSILE® (see Figure 2.6) notes to be related to more than one other node and offers several views on the relations between nodes, like 'history', 'refers-to' and 'referred-by'.

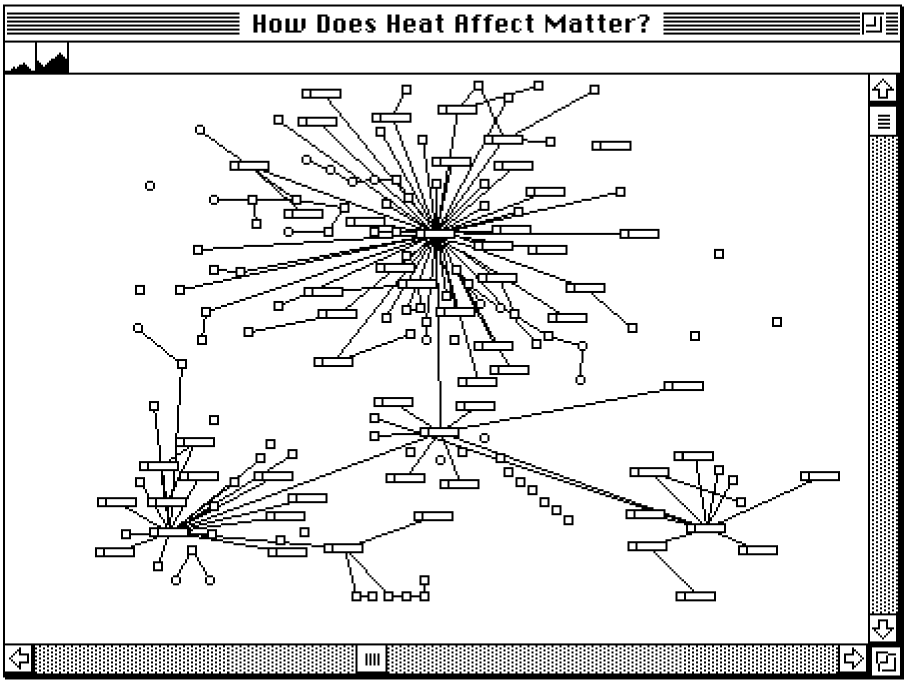

Figure 2.6. Enlarged CSILE® Knowledge Map from Hewitt, Scardamalia, and Webb (1997).

Conclusion

CSCL environments are designed to be used by learners to increase their knowledge by engaging in collaborative problem solving or argumentation. Because they are meant for learners / novices, they are easy to (learn to) use and often have to offer simplified views on domains. Those environments that offer graphical argumentation were primarily developed to support novices in scientific inquiry and explanation. They offer objects and relations specific to these purposes (TC3® diagrammer, that offers more general argumentation facilities is an exception here). Their orientation towards novices is reflected in the representational services that they offer, namely the learners can use a basic and limited set of objects and relations with which typical 'why'- questions can be explored. Services to maintain different problem perspectives or support the use of (partially) shared representations, as well as more advanced services to help maintain coherence and detect conflicts are rarely found. Most of the systems support only one way of communication and representation in collaborative problem solving and in no way support learners moving back and forth between different styles of communication and representation in different phases of problem solving.

Considering all this, it is no wonder that CSCL environments offer only a glimpse of what is needed to support collaborative solving of social science problems with the help of CSAV. Two things are missing. First, the representation needs to be able to represent more than only the 'Why'-side of as problem, the analysis. It needs to support the representation (and discussion) of such topics as constraints and criteria, possible actions and likelihood of success. Second, the representation cannot be limited to a particular domain, as is often the case for CSCL environments. Different perspectives, stakeholders with particular viewpoints, are involved in 'social science' problems. In other words, the representation needs to be able to deal with broader domains, covering topics dealing with constraints, criteria and 'How'-questions. Before drawing any final conclusions, we discuss a number of CSAV environments for deliberations dealing with typical 'How'-questions.

Representation and support for argumentation in design-based environments

A series of computer-based tools have been developed within the planning and design tradition to support the solving of wicked problems or to record design rationale. Design rationale is a systematic approach to lay out the reasons for and the reasoning behind the decisions that led to the design of an artifact (Buckingham Shum & Hammond, 1994; Carroll & Moran, 1991). Tools to support expression of design rationale allow or force users to state alternatives, evaluate these against criteria and constraints, and finally decide on alternatives that best reach the goals given the constraints in which the designer has to operate. Thus, these systems represent a type of argumentation dealing with several of the topics that were not, or not sufficiently, covered in the CSCL environments discussed in the previous section. We will not discuss the research in design rationale and the systems to record design rational at length, but rather concentrate on two (representational) notations frequently used in the field, since these notation are at the roots of most systems and they provide us with the objects and relations for argumentation dealing with the 'How'-questions.

The IBIS notation

Most design rationale systems are rooted in Issue-Based Information Systems (IBIS), introduced by Kunz and Rittel (1970) to support argumentative problem solving. IBIS invites the participants to raise questions (*Issues*), formulate alternative answers (*Positions*), to which pros and cons (*Arguments*) may be attached. Decisions as to which answers are accepted or rejected are called *Resolutions*. IBIS uses several types of links between the nodes, such as responds-to, questions, supports, specializes, and generalizes. IBIS is not a freeform method. It has a small grammar that specifies how objects may be related and the environments enforce correct use of the objects and relations that help keep the representation coherent. Examples of computer implementations that stick close to the original notation are gIBIS® (Conklin & Begeman, 1988), QuestMap® (Figure 2.8) and Mifflin® (Selvin, 2003).

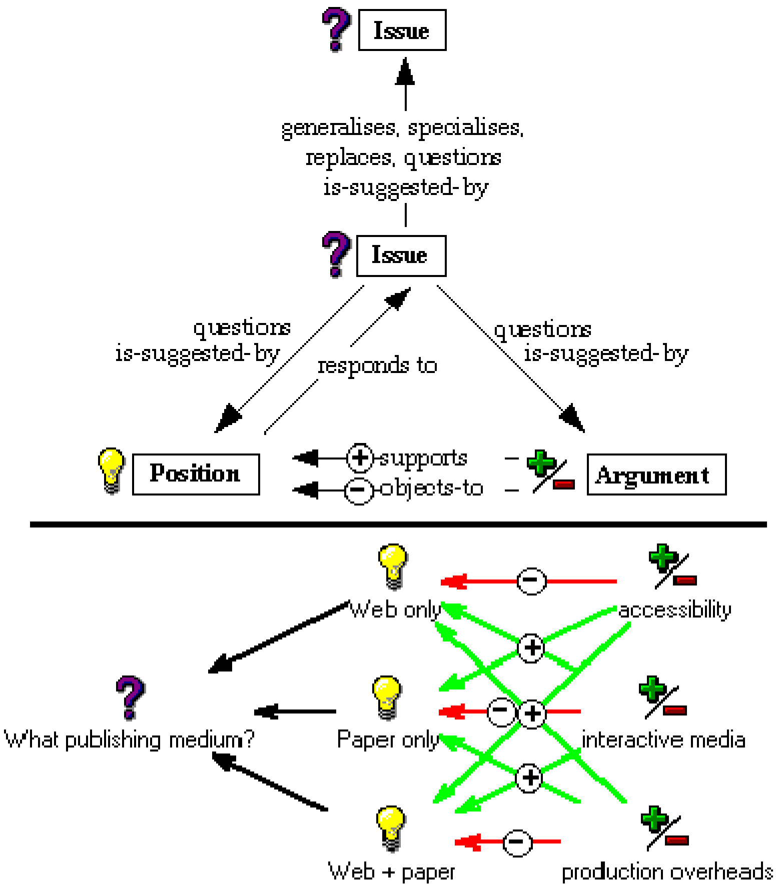

Figure 2.7. The graphical IBIS (gIBIS®) notation and an example, showing how Issues, Positions and Arguments are combined to cumulatively build graphical argument spaces (Buckingham Shum et al., 1997).

IBIS can represent deliberations on a starting question, like 'How can we reduce school drop-out?' This question is the start *Issue*, to which *specializing questions* can be attached. One might for instance start a discussion on problem constraints by raising the *Issue* 'How much time and budget do we have?' or raise the analytical *Issue* 'What causes drop-out?' Hypotheses regarding the causes, formulated as *Positions* and *Arguments*, can be added to support or refute these hypotheses. *Issues* may be added to raise questions on the validity of certain *Arguments*, other questions asking for facts can be added and in a similar way an *Issue* object can be formulated to start a discussion of the constraints or criteria that any solution has to meet.

IBIS implementations are less dedicated in their ontology than the CSCL-based systems discussed: the IBIS-notation is non-committal to specific uses, which has the advantage that it can serve to represent different views of a problem in the initial stages of problem solving. Not surprisingly, IBIS-based systems are strongest where CSCL-based systems are weakest. The IBIS style supports issue-based communication, brainstorming and other activities found in the orientation state more easily than the CSCL-based systems. But there are drawbacks, especially in educational settings. The issue structure is weak and the representational notation does not force important categorical choices (for instance between alternatives, constraints, and interventions). IBIS is a method that supports initial problem structuring, but offers no specific means for formulation of solutions (specializations of IBIS were proposed to offer more specific support in design which we will not discuss here). In the Compendium methodology (Selvin, 2003) IBIS is used in conjunction with formal modeling. Here the facilitator can introduce pre-defined data structures or even create data structures that offer a more accurate representation than standard IBIS can. Note that this implies that partially shared representations may be used. None of the IBIS implementations we know of has other computational means to ensure coherence, detect conflicts or manage plausibility.

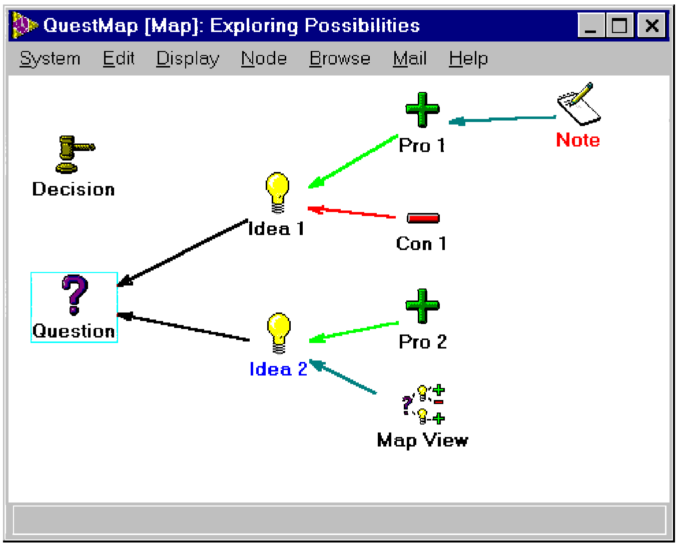

Figure 2.8. QuestMap® IBIS skin.

Decision Representation Language

Decision Representation Language (DRL) is a notation developed to support qualitative decision making (Lee, 1990). The most important objects in DRL are *Alternatives*, *Goals* and *Claims*. *Alternatives* are the options between which a decision has to be made. *Goals* specify the desirable properties of an option. The top-level node of the goal hierarchy is a *Decision Problem*. *Claims* in DRL are used to represent arguments that are relevant to a decision between alternatives, and *all relations* in DRL are specializations of the claim object. A claim is related to another claim by a *support*, *denies* or *qualifies* relation. In other words, all objects and relations in DRL are refutable. Figure 2.9 presents the main elements of the DRL notation.

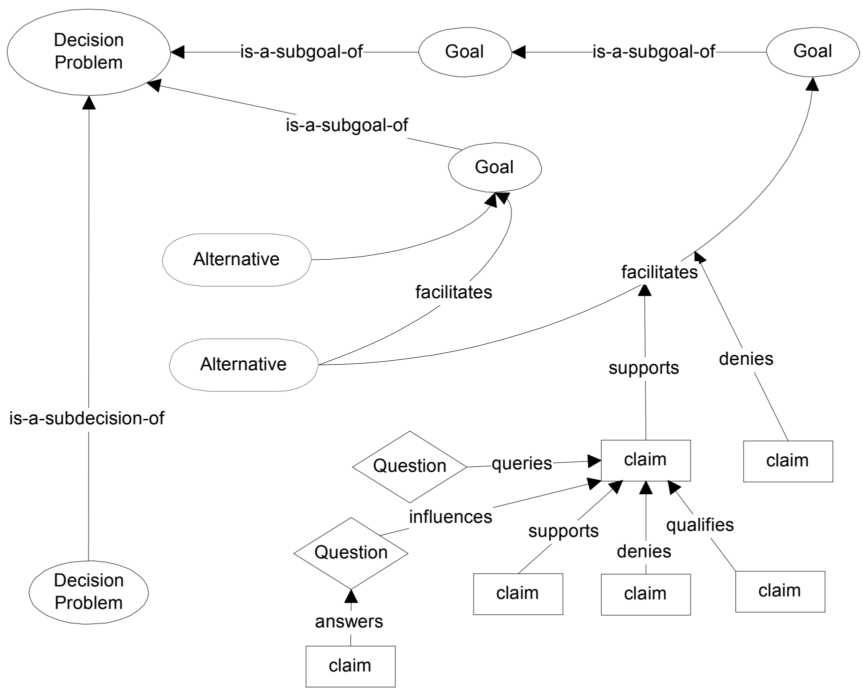

Figure 2.9. Part of the DRL notation adapted from Lee, 1990 p. 112*)*.

SIBYL®, a prototype of a qualitative decision management system (Lee, 1990) is a computer implementation of DRL that offers a number of guidance and support services including *plausibility management*. The system manages plausibility by updating the a-priori plausibility of a claim once new claims are added that support, qualify or deny the original one. Plausibility management is a special form of dependency management that maintains a consistent state in the knowledge base when changes are introduced. In SIBYL® users can be associated with a claim, and the number of associated users can be used as a parameter in the procedures that update plausibility measures. Consider an example of SIBYL's services applied to the arena of scientific inquiry. Here one might qualify the claim that *data* D supports *hypothesis* H by pointing out that D is measured using an unreliable instrument. If the plausibility of this statement surpasses a certain threshold, plausibility management will set the plausibility of the claim *supports* (D,H) to zero. On the larger scale of dependency management the effects may be further propagated.

SIBYL® is a highly advanced implementation that requires strict use of the formalism of DRL. We do not intend to use DRL in such a way. The notation has a number of advantages that also in informal use allows users to express more than the notations encountered in CSCL or in IBIS. In DRL, for instance, one can discuss whether a piece of data is really supporting a hypothesis and one can even refute the data. Belvédère®, in contrast, treats 'data' more or less as givens, that cannot be refuted by other data, because only evidential relations are supported (Chryssafidou, 1999). A first approximation of a DRL-based representational notation is presented in the next section of this chapter.

A representational notation to support social science problem solving

One of the goals of the present research is to develop a representational notation to support collaborative solving of social science problems. In the previous sections we saw how CSCL and design-based environments using external representations each support particular types of problem solving. Most CSCL environments concentrate on the 'Why'-aspects of problems, allowing their users to express claims about causes (hypotheses) and bring forward supporting or contradicting evidence (data). Design-based environment put more emphasis on the 'How'-side. DRL for example supports reasoning about a number of alternative solutions to solve a problem. A combination of the representational notation found in the literature is needed to cover the full area of social science problem solving.

Here, we present a first attempt at a representational notation to support collaborative solving of social science problems. The notation aims at allowing representation of the argumentation *underlying* the analysis and solution of the problem. It does not contain the kind of meta-cognitive scaffolds found in CSILE® or Knowledge Forum®. We are specifying a representational notation to capture claims and arguments about the analysis and solutions of problems. Whether all objects and relations of this notation are used and whether the notation as a whole is usable needs to be addressed as a separate issue (and one that we address in chapters 4 and 5).

The notation is based on the DRL notation just presented. According to DRL all objects are specializations of a *claim* object. Thus, if one introduces a hypothesis (or cause) X, this in fact means that a claim is made that X is a cause. All claims can be challenged and they all are open to debate and refutable. The notation uses a number of standard relations of DRL such as 'supports', 'denies', 'qualifies', 'influences', 'queries', 'answers' and 'facilitates'. We will not discuss these in detail. Remember, that these relations are also claims. So, if a 'data object' is connected to a particular hypothesis by means of a 'support'-relation, then a *claim* is added that the data supports the hypothesis. This claim may be challenged by other claims, for instance one that asserts that the data is not measured in a valid way. An impression of the notation is presented in Figure 2.10.

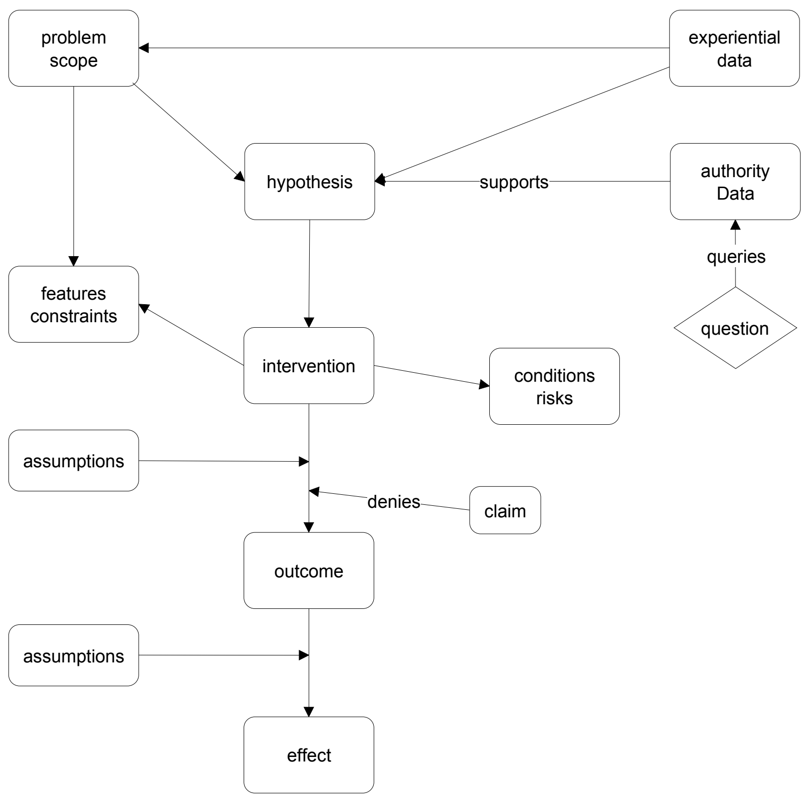

Figure 2.10. DRL-based representational notation.

We used concepts from the CSCL inquiry environments as well as from the design-based systems. From the inquiry environments objects such as *hypothesis* and *data* are used to make claims about potential causes and add data to support these claims. In order to allow differentiation between personal experience and data sources open for inspection, we have made a further subdivision between experiential and authority data. Neither the CSCL-inquiry environments, nor the design-based systems offer means to express constraints to the problem or other statements indicating the scope of the problem. Experts, however, typically spend a substantial amount of time in getting these aspects of the problems clear (Voss, Tyler, et al., 1983). For that reason *constraints and features* as well as *problem scope* were added to the objects of the representational notation.

The expert solution to a social science problem is, according to Voss, Tylor, et al. (1983), often abstract ('improve the infrastructure') and needs further detailing before it is a executable plan. Such a plan can become a complex set of interventions grouped into a project or a program. In order to support making statements (claims) about such a plan the representational notation contains concepts from Logical Framework Analysis (Sartorius, 1996) and evaluation theory. Logical Framework Analysis (LFA) is a set of tools and methods for analyzing the causes of a social problem and formulating interventions that, through a chain of effects, is thought to solve the problem. LFA emphasizes the importance of the 'project logic' that should exist in for instance the causal relations between interventions, short term and long-term effects. Objects in the notion such as i*nterventions*, their *short-term outcomes* and *long-term effects* as well as (pre)conditions and risks are based on LFA. LFA does not encourage or force participants to explicate the 'theory' behind the project as theory-driven evaluation demands (Chen, 1990; Rossi, Freeman, & Lipsey, 1999). To encourage statements about the program logic the object *assumptions* was added to express the reasons (empirical or theoretical) why the intervention may be expected to have the expected effects on the target population. By making different groupings of interventions, outcomes, effects and assumptions alternative plans can be created. A simple example of the application of this notation is presented in Figure 2.11.

The representational notation can be considered an eclectic blend of several other notations. Its combination of characteristics make it suitable to represent the argumentation underlying the analysis, solution and its evaluation of social science problems. First, there is explicit support for problem orientation: objects are made available to express the *scope* of the problem (only highway accidents are considered) as well as its *constraints* (no changes to infrastructure). Preliminary ideas about the *causes* (high speed) can be formulated as hypotheses. When causes are further explored, authority (incidents statistics) as well as experiential *data* can be added to substantiate or weaken statements about causes. A *question* (are there statistics for highways?) can be added to focus attention to a pending issue. The formulation of a solution consists of an *intervention* (obligatory speed governor) accompanied by claims about the short-term and long-term *effects* (slower cars, less fatalities). The intervention is tested (evaluated) against stated constraints and, more or less as a rebuttal, (pre)*conditions* and *risks* that accompany the interventions can be described (speed governor will be rigged). *Assumptions*, assumed intervening mechanisms of the program's theory or important *ceteris paribus* clauses, can be stated (driving habits will remain unchanged; impact speed will be reduced). Experiential as well as authority *data* can be added to support or weaken the claim that the solution will work, aiding evaluation.

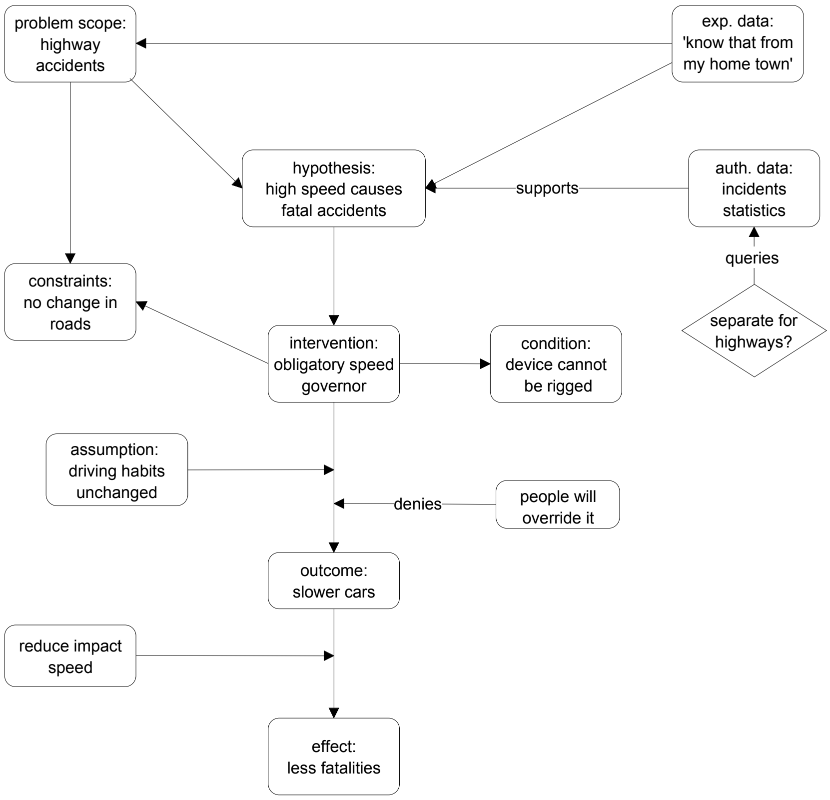

Figure 2.11. Application of the representational notation to a fictitious example.

Conclusions

To get a clearer understanding of how external representations can be used to support collaborative solving of social science problems, we reviewed research in the psychological as well as the design approach to these problems. We tried to derive a number of demands from the description of problem solving activities and translate these in terms of requirements on services to be performed by a CSAV environment. We then compared these requirements to actual services found in some well-known CSCL environments. Finally, we concentrated on the core issue of the representational notation and discussed two of these notations that emerged from the design tradition. An implementation of the Decision Representation Language (DRL) was briefly discussed and shown to offer at least the same level of guidance and support as found in the CSCL environments.

The core element of a system using external representations is the representational notation: the set of objects and relations and the rules of how to combine them. We derived such a representational notation from a number of sources. As stated above, we need to find out whether the notation as a whole is usable. The representational notation developed here is rich and rather complex. Elsewhere (Van Bruggen & Dekeyser, 2001; Van Bruggen, Kirschner, & Jochems, 2002) we have argued that rich representations may effectively put an additional cognitive burden on the learner, rather than offering off-loading effects (Duffy & Cunningham, 1996). Two questions have to be answered. *First*, are learners supported by such a complex representations when they try to derive adequate solutions? Or are simpler representations sufficient? *Second*, how usable is this complex representation for learners? In chapters 4 and 5 we address the second question by using the representational notation to analyze the dialogues of students that are solving problems. The assumption being that if students already use concepts contained in the notation, the notation is probably easier to learn and apply.

Before plunging into an implementation of the rich representational notation, we will consider whether more modest alternatives might work as well. In the next chapter we report on an experiment in an existing CSCL environment. A slightly adapted representational notation was introduced to find out whether that adapted notation supported collaborative solving of a social science problem.

3 Using Belvédère for social science problem solving[4]

The experiment discussed in this chapter is concerned with the use of computer-supported argument visualization (CSAV) in collaborative solving of social science problems (Voss, Greene, et al., 1983; Voss, Perkins, & Segal, 1991). These are problems where the problem solvers need to rely on informal reasoning to reach a solution that seems *likely* to solve the problem and is *acceptable* to the problem solvers. We used an existing tool (Belvédère®) for collaborative inquiry, but added slightly different meanings to its symbols to not only represent the analytical or *Why*-side (Why did X occur?), but also the more practical *How*-side (How can we reduce X?).

Belvédère® was developed to support scientific inquiry. Its representational notation corresponds to that domain: there are hypotheses (H) and data (D) and the relations between objects are evidential. One can express that D is consistent with H. One cannot express that D refutes H, or that D is a likely event. In our case, one can do the 'Why'-side of the problem using the representational notation of Belvédère®, but there are no specialized means to express the 'How'-side. Notwithstanding its dedicated nature, Belvédère® has been used as a more general tool to express argumentation (e.g., Veerman, 2000). One may ask whether the 'lack of fit' between the task of solving a social science problem on the one hand, and the ontology and specificity of the representation on the other has any notable effect on problem solving. In this study we are interested in whether changes in the ontology and specificity of the objects in the external representation have any notable effects on the Belvédère® diagrams and the solutions that are proposed as a result of solving a social science problem. We restrict ourselves here to the representational notation and its representational guidance and do not address changes in the problem solving process or process guidance.

The core hypothesis of this study is that (1) by incorporating objects in the representational notation of Belvédère® that are pertinent to 'how'-questions and (2) by defining the meaning of other objects in such a way that 'how'-aspects are included, a guidance effect is triggered that will bias the learner towards the 'how'-questions rather than to merely analyzing the problem. There are a number of predictions related to this hypothesis. First, we do not expect that the richer ontology and higher specificity will lead to an easier processing of the material. On the contrary, these changes to the notation will require more processing, because participants have to make more categorical choices. This may affect such variables as time needed to process the material, the mental effort required to process the material et cetera. We do not address all of these effects, but concentrate on the effects on the external representations prepared by our participants, and their retention of the background material and their recommendations. We expect that a richer ontology and a higher specificity will lead to richer and more connected diagrammatic representations because participants need more objects and relations to represent the background material and their recommendations. As stated above, we expect higher specificity to require more processing and we expect this to be reflected in higher retention rates of the material and the results of their work, their recommended solutions to a social science problem.

---

[4] We would like to express our gratitude to prof. dr. Martin Valcke, University of Ghent, for making this experiment possible and to our colleague Henk Münstermann for suggesting the topic and making available material on special and inclusive education. Parts of this chapter were presented at EARLI 2001 (Van Bruggen & Dekeyser, 2001)

Design and methods

Participants

Second year Education students (*N*=95, 86 female, 9 male) studying at the University of Ghent, Belgium participated in this study. The participants were randomly assigned to pairs (43) or occasionally to triplets (3), to work together on the experimental task. Pairs or triplets were randomly assigned to a High or Low Detail condition (see below).

Design

Participants, in pairs or triplets, worked on the solution of a social science problem, namely how to swiftly implement inclusive education in Belgium. Inclusive education is defined as the inclusion of pupils with (severe) handicaps in 'normal' schools. The alternative is maintaining special education schools or classes for these pupils. The pairs and triplets worked on solving the problem with the aid of a modified version of Belvédère® version 2.

The experiment had a High Detail and a Low Detail condition. In the Low Detail condition participants were instructed to only use the objects *Hypothesis* - to represent assumptions - and *Data* to represent facts in their argument visualizations. In the High Detail condition participants were instructed to use the Belvédère® objects as follows: "*Hypothesis*: assumptions, in particular assumed problems for implementation"; "*Principle*: solutions to problems", "*Data*: facts about the assumptions or the solutions", "*Unspecified*: use this object to indicate where you need additional information". Thus the High Detail condition had more objects (i.e., a richer ontology) as well as objects with a more specific meaning. An example of the use of this richer ontology is presented in the Belvédère® diagram in Figure 3.1.

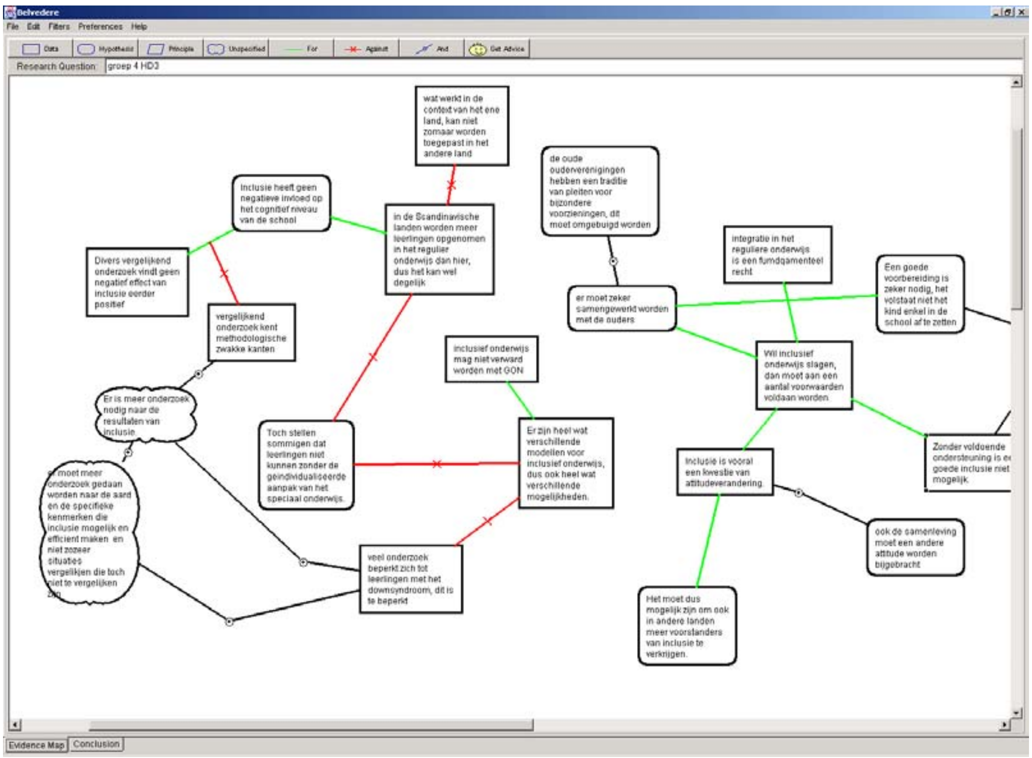

Figure 3.1. Belvédère® diagram from a High Detail group.

Procedures and Materials

Each group (pair or triplet) participated in one of the six sessions in which the data were collected. Each session lasted two hours. In these sessions the participants, after receiving instructions and filling out questionnaires, started to work using Belvédère®. Assistants were available to help students get acquainted with the Belvédère® environment. The second set of data was collected in a post session which took place five to seven weeks after the experimental session.

The participants were instructed to compile recommendations on a rapid implementation of inclusive education in Belgium. We selected this subject because we assumed that Education students would find the topic interesting and we expected them to have some general knowledge of inclusive education, but not on the level of specific methods used in special education, specific implementation strategies et cetera. Participants were instructed that they should take into account criticism of inclusive education in their recommendations. They were given background material covering diverse aspects of inclusive education. Some of the material was considered not to be essential to compiling recommendations. It was added to ensure that participants would have to make a selection from the material for further consideration.

The participants were instructed to use the Belvédère® environment to process the background material and to formulate their recommendations. The participants each worked on a computer that was running Belvédère® in 'client mode'. In this mode the computers are connected over the Internet to a central Belvédère® server, which updates the client machines and stores the data (diagram, chat and conclusions). For the experiment we used a Belvédère® server located at the Open University of the Netherlands. Because the Internet connections to the Belvédère® server were not available for two of the sessions due to technical problems, we instructed the students in those sessions to work side-by-side at a single computer that ran a stand-alone version of Belvédère®. This affected ten groups, with in total twenty students. We decided to keep these students within the analyses since the way they had to cooperate – using a shared computer – is the modus operandi in most Belvédère® research. We contrast this collaboration mode (shared mode) with one in which students each operated their own computer (separate mode).

Data collected

A diverse set of data was collected during the experimental sessions and one post-session. First, a simple prior knowledge test was used to determine whether students had specific prior knowledge with respect to inclusive education. Considering the task (compile recommendations on the implementation of inclusive education) we do not expect general knowledge to be helpful in completing the task. We asked the participants to rate, on a 4-point scale (1 = no knowledge of the concept, 4 = well known concept), their familiarity with general concepts related to special and inclusive education (e.g., Down's Syndrome) as well as with concepts related to specific policies, projects and methods.

Second, we gathered data on the experience the participants had with different types of computer applications. Although Belvédère® is very intuitive to use, it still may be hard to use for the complete computer novice or for someone without any typing skills. We therefore asked the participants to rate their experience with different types of computer applications (e.g., word processing, mail, chat, discussion groups) on a 4-point scale from no experience (=1) to very experienced (=4).

A third set of data was collected at the end of the session. Participants used rating scales to describe how relevant, new and difficult the background material was, and whether the time allotted was sufficient to complete the task.

The core of the data was collected on the outcomes of the problem solving process, namely the recommendations and the Belvédère® diagrams that the participants compiled. The Belvédère® diagrams that the students created during the sessions were recovered from the Belvédère® server or from the stand-alone systems (sessions 1 and 4). Participants entered their recommendations into the Belvédère® conclusion window or put them down in writing in the sessions where stand-alone systems were used (1 and 4). Data loss due to data-transport problems restricts us to reporting on only a subset of the diagrams and recommendations entered. For similar reasons we cannot report on the Belvédère® chats. Obviously there are no chat records for the two stand-alone sessions, but the data of the other sessions were affected as well by data transport problems and subsequent hardware failures that left only pieces of the chats.

The last set of data was collected during one post-session that came five to seven weeks after the experimental sessions. During this session students answered three factual questions on the background material and produced a written recall of the recommendations, problems and solutions that they had compiled during the experimental session.

Results

Prior knowledge of special and inclusive education

Prior knowledge of special and inclusive education was measured using a questionnaire in which participants rated their knowledge of a number of general concepts (1, 2, 6, 9, 12), policies and projects on implementing integrated or inclusive education (4, 5, 8, 11) and pedagogical methods (3, 7, 10) found in special and inclusive education. The main statistics of the items in the prior knowledge scale are reported in Table 3.1.

Table 3.1
Statistics of the Items on Prior Knowledge of Special and Inclusive Education (N=95)

	Concept	*M*	*Md*	*SD*
1	Special Education	3.31	3	.64
2	Down's Syndrome (n=94)	3.25	3	.71
3	Feuerstein method	1.39	1	.59
4	Normally different project	1.20	1	.47
5	Integrated Education	3.01	3	.89
6	Inclusive Education (n=94)	3.65	4	.50
7	Instrumental Enrichment	1.09	1	.29
8	Mentor project	1.37	1	.62
9	Psycho-Medical Social Center	3.67	4	.52
10	Gesture-supported speech	1.49	1	.91
11	Salamanca declaration	1.33	1	.57
12	Learning handicaps	1.06	1	.35

Note. 1 = unknown, 4 = well known

Table 3.1 shows that students reported greater familiarity with the concepts related to inclusive and special education in general and with the Integrated Education initiative than with the specific policies, projects and methods to foster inclusive education. Since the task did not require the participants to come up with implementation plans specific to certain learning or behavioral handicaps, but rather for more general implementation plans for inclusive education, we decided to use all the items in an index for prior knowledge. The index was calculated by totaling the scores of the items and dividing them by the number of valid responses (M = 2.15, SD=.28)[5]. The reliability of this index in terms of internal consistency was found to be acceptable (α = .67).

Experience with different types of ICT applications

Prior experience with ICT was measured by having participants rate their experience with different *types* of computer applications. The results are summarized in Table 3.2.

[5] Although the calculations and analyses reported here and elsewhere in this chapter may require – formally speaking – interval level measurement of the ratings, we have decided to treat the ratings, according to statistical practice, as interval data (see Lord, F.M., & Novick, M.R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley, p. 22).

Table 3.2
Statistics for the Items on ICT Prior Knowledge (N=95)

Type of Application	*M*	*Md*	*SD*
Text processing	3.19	3	.57
Spreadsheets	1.45	1	.63
Presentation software (*n*=94)	1.72	2	.72
Database	1.45	1	.60
Drawing	1.66	2	.69
Email (*n*=94)	3.44	4	.66
Web browsers	1.72	1	.88
Discussion groups	2.53	3	.93
Chatting (*n*=94)	2.41	2	1.04
MUDS and MOOs	1.06	1	.24
Collaborative software	1.24	1	.56

Note. 1 = no experience, 4 = lot of experience

The summary statistics of Table 3.2 show that the participants are most experienced with text processing, email, discussion groups and chatting. All this suggests that there is no reason to assume that lack of computer handling skills could bias the results of the experiment. For further analyses an index for prior experience on the mean of the valid scores was prepared. This index has a mean of 1.99 and an *SD* of .36. The reliability in terms of internal consistency ($\alpha = .71$) is acceptable for the purposes of the research.

Ratings of the background material and available time

We asked participants to rate (in percentages) how much of the background material was *relevant* and how much of it was *new*. We further asked them to rate the *difficulty* of the material on a 5-point scale ranging from "(1) very difficult" to "(5) very easy". Finally participants used a 5-point scale ranging from "(1) far too little time" to "(5) far too much time" to rate whether there was sufficient time allotted to complete the task. Summary statistics of these measures can be found in Table 3.3. The table contains a breakdown of the data according to Detail condition as well as one according to Collaboration condition.

Table 3.3
Ratings of the Background Material and the Time Available for the Task per Detail and Collaboration Conditions

	Detail		Collaboration		
	Low	High	Separate	Shared	Total
	Percentage of material considered relevant				
M	53.44	59.69	56.32	57.69	56.70
Md	50	60	60	60	60
SD	22.07	18.80	21.15	19.30	20.56
N	45	49	68	26	94[a]
	Percentage of material considered new				
M	53.78	53.64	51.84	58.65	53.71
Md	50	53.5	50	55	50
SD	16.89	18.54	17.78	16.77	17.69
N	45	50	69	26	95
	Difficulty of the material				
M	3.07	3.12	3.07	3.15	3.09
Md	3	3	3	3	3
SD	.62	.66	.65	.61	.64
N	45	50	69	26	95
	Time available for the task				
M	2.24	2.60	2.32	2.73	2.43
Md	2	2	2	2	2
SD	.98	1.03	.93	1.19	1.02
N	45	50	69	26	95

[a] One student did not answer this question

Most of the material (53 – 60 %) was considered relevant and the majority (54% overall) of the material was rated as new by the participants. The difficulty of the material was rated as average ($M = 3.1$ on a 5-point scale). Finally, the time available for the task is rated as being limited ($M = 2.4$ on a 5-point scale). There is no statistically significant relation between these ratings and the prior knowledge index (r varied from -.08 to .13).

We expected that higher specificity of the notational system would make the material harder to process and that this would result in participants perceiving the material as more difficult and the time available as less adequate. However, students in the High Detail condition rated the material as easier and the available time as more adequate. Only the ratings of the relevancy of the material by students in the High Detail condition were higher than those in the Low Detail condition. We explored the latter contrast and the ones in the Separate-Shared conditions using analysis of variance, that is, assuming interval scale measurement. No contrast was statistically significant. Thus, contrary to our expectations, specificity (and ontology) had neither an influence on the perception of the relevancy, difficulty or novelty of the material, nor on the evaluation of the time available for the task.

Recommendations

The participants' main task was to draw up recommendations on how to implement inclusive education. The recommendations and the diagrams that they created reaching those recommendations are the joint products of two or three participants working together. Therefore, our analysis uses the group as observational unit. Data are available for only 28 groups (59 participants) due to loss of data on the server. The recommendations were copied from the Belvédère® server or, if they were handwritten, typed out after which they were scored by one trained and experienced coder. This coder had already done the coding of the recall of the recommendations (reported in Table 3.11) where the same coding protocol was used. Considering the high correlations that were achieved at that coding, we considered the coding procedure to be reliable enough to have only this one experienced coder code the data. The protocol asked the coder to differentiate between recommendations and problems and their respective elaborations. The basic statistics of the results are shown in Table 3.4. The most remarkable result is the low mean for problems identified for implementation of inclusive education. Although they were instructed to identify problems for the implementations, participants mention and elaborate only a few problems (on average less than one), while more than five recommendations on average are given.

Table 3.4
Statistics for the Number of Recommendations and Problems and their Elaborations of the Groups (n=28)

	M	*SD*
Recommendations	5.29	3.79
Elaborations	1.29	1.12
Problems	.79	1.34
Elaborations	.21	.50

We predicted that more recommendations and more elaborations would be made in the High Detail condition. Table 3.5 reports the means and standard deviations for the number of recommendations and problems and their elaborations. Unfortunately, there is a lack of balance due to technical problems; we have 9 Low Detail groups versus 19 High Detail groups. The differences between the means of the conditions are small and none of the differences approaches statistical significance.

Table 3.5
Recommendations and Problems for Low and High Detail Conditions

	Detail Condition	
	Low (n = 9)	High (n = 19)
	Recommendations	
M	5.78	5.05
SD	4.84	3.31
	Elaborations of recommendations	
M	1.11	1.37
SD	1.36	1.01
	Problems	
M	.67	.84
SD	1.12	1.46
	Elaborations of problems	
M	.11	.26
SD	.33	.56

Table 3.6
Recommendations and Problems and their Elaborations broken down by the Conditions of the Experiment

	Collaboration mode			
	Separate		Shared	
Detail	Low (n = 4)	High (n = 11)	Low (n = 5)	High (n = 8)
	Recommendations			
M	4.50	3.91	6.80	6.63
SD	4.20	1.97	5.54	4.21
	Elaborations of recommendations			
M	2.00	1.36	.40	1.38
SD	1.41	1.03	.89	1.06
	Problems			
M	.75	1.09	.60	.50
SD	.96	1.70	1.34	1.07
	Elaborations of problems			
M	.00	.27	.20	.25
SD	.00	.47	.45	.71

A multivariate analysis of variance showed no statistical effects for any of the main or interaction factors, although the main effect of Collaboration Mode nearly reached the 5% limit (Λ = .663, $F_{(4,21)}$ = 2.67, p = .06). Subsequent univariate analyses revealed no significant differences between the means, but hinted to the low number of elaborations of the recommendations in the low detail-shared condition as causing the 'near-effect'. All in all, the data do not confirm the prediction that the High Detail condition would lead to formulation of more recommendations and problems or to more elaborations.

Belvédère® diagrams

There were 43 Belvédère® diagrams available for analysis across the different conditions (see Table 3.7).

Table 3.7
Number of Diagrams Obtained in the Conditions of the Experiment

	Detail Condition	
Collaboration mode	Low	High
Separate	15	15
Shared	5	8

Note. There were 89 participants in the 43 groups

The content of the diagrams varied widely. Whereas some groups entered several concepts and relations to diagram substantial portions of their analysis of the background material, others produced diagrams with very few concepts and relations. For the scoring of the diagrams, it is important to realize that the diagrams here are not used to represent a scientific argumentation. Measures that are based on a balance between pro- and contra arguments (e.g. as used by Veerman, 2000) are therefore not satisfactory for our purpose. Moreover, since the groups used different sets of objects, we cannot use measures that are dependent on the existence of one underlying object set. We have therefore adopted a very simple scheme and rated the diagrams in terms of the richness and connectivity of their content. Richness is defined as the total number of objects and relations in the diagram and connectivity is measured as the number of relations per object in the diagram[6].
The main characteristics of the Belvédère® diagrams are reported in Table 3.8. Looking at the Detail conditions, we see that students in general did what they were instructed to do. In the Low Detail group the diagrams contain fewer objects of the principle type. The same holds true for the number of unspecified objects (the number of the Low Detail group reported here is inflated by one diagram that has 17 objects of type unspecified). Other outcomes are less obvious. Diagrams in the Low Detail condition show higher richness than those made in the High Detail condition. The same difference exists between the Separate and Shared collaboration mode.

[6] Strictly speaking, the maximum number of relations is the combination ${}^{N}C_{2}$ where N is the number of objects. This correction makes sense if the number of objects is very small. For somewhat larger numbers the correction looses sense both computational - the combination grows rapidly – and semantically as the correction implies a comparison to a representation where all possible connections between all objects are made.

Table 3.8
Overall means of the characteristics of the Belvédère® diagrams for the detail and collaboration condition

	Detail		Collaboration mode	
	Low (n = 20)	High (n = 23)	Separate (n = 30)	Shared (n = 13)
Objects	20.45	16.30	19.90	14.38
Relations	18.40	15.52	18.93	12.08
Richness (Σ objects + relations)	38.85	31.83	38.83	26.46
Connectivity (relations per object)	.88	.90	.93	.80
Disconnected objects	.95	.48	.57	1.00
Data objects	12.05	6.78	9.47	8.69
Hypotheses	6.65	5.30	7.07	3.31
Principles	.70	2.91	1.77	2.15
Unspecified	1.05	1.30	1.60	.23
For relations	9.60	8.74	10.07	7.00
Against relations	4.90	3.78	4.90	2.92
AND relations	3.90	3.00	3.97	2.15

The High Detail groups produced diagrams with less richness than the Low Detail groups. The results for connectivity are almost the same for both conditions. Thus with the exception of a slightly higher number of relations per object for the High Detail condition all data point to an opposite direction than the one predicted: higher specificity and richer ontology did not result in richer diagrams with higher connectivity. The evidence suggests that exactly the reverse is closer to the truth. Inspection of the collaboration mode data in Table 3.8 shows that the differences between the means here are larger. This is reported in more detail in Table 3.9.

Table 3.9
Means of Diagram Characteristics and Number of Groups for the Two Conditions

	Collaboration mode		
	Separate	Shared	Total
Detail		Richness	
Low	41.20	31.80	38.85
	(n = 15)[a]	(n = 5))[a]	(n = 20)
High	36.47	23.13	31.83
	(n = 15))[a]	(n = 8))[a]	(n = 23)
Total	38.83	26.46	35.09
		Connectivity	
Low	.92	.78	.88
High	.95	.82	.90
Total	.93	.80	.89

[a] The number of observations for connectivity are equal to the data presented here and are not repeated.

The data in the table suggest that both Collaboration Mode and Detail have an effect on the richness and connectivity of the diagrams. Collaborating at separate workstations using a representational notation with low specificity appears to lead to diagrams with the highest richness, while working at a shared workstation with high detail leads to diagrams with the lowest richness. A multivariate analysis with richness and connectivity as dependent variables and Detail and Collaboration Mode as factors showed no overall significant effect for Detail ($\Lambda = .925$, $F_{(2,38)} = 1.54$, $p = .23$), but the effect for Collaboration Mode ($\Lambda = .865$, $F_{(2,38)} = 2.965$, $p = .06$) was such that we explored it further. Univariate tests of the between-subjects effects revealed a statistical significance for the Collaboration Mode factor for richness ($F_{(1,39)} = 4.229$, $p = .05$) and connectivity ($F_{(1,39)} = 4.83$, $p = .03$). Apparently, collaboration mode affects the richness of the diagrams. These results, however, have to be interpreted with caution: the Collaboration Mode condition was introduced more or less by accident and it resulted in an unbalanced design. The univariate analyses exhaust the degrees of freedom available for independent testing and the underlying data show large differences in the number of observations per cell with substantial differences in standard deviations between the cells as well. Replication of this study is therefore necessary.

Retention of recommendations

During the post-session we asked the participants ($n = 75$) to recall and write down the recommendations they had compiled at the experimental session, the problems they had expected and the solutions to those problems. These written recalls were typed out and scored according to a detailed scoring protocol. Separate scores were given for the number of recommendations, number of identified problems and number of solutions in the written recalls. For each of these categories the number of different elaborations that the participants added to these topics was counted. Note that this is a coding scheme that does not take into account the accuracy of the retention, but is solely based on the recall written down at the post-session. Thus the imagination of the participants may have led to recommendations that are beyond the ones they produced during the experimental session. On the other hand, this is a bit of a moot point, because we do not know what participants discussed during the experimental session.

The data were coded independently by a trained coder who was experienced in coding problem solving protocols and the author. Inspection of the raw scores showed that there were major discrepancies in 6 of the first 61 cases, that is, in 6 of the 366 scores given. These discrepancies were checked with the coder and led to a recoding of three items. The codes of the other items were discussed, but no changes were made. All analyses are based on the (corrected) codes of the trained coder only. The correlation between the coders is reported in Table 3.10.

Table 3.10
Number of Participants who provided Data and the Pearson Correlations Between Coders on the Counts of Recommendations, Problems and Solutions

	n [a]	Number	Elaborations	Total
Recommendations	61	.87	.75	.89
Problems	66	.87	.50	.82
Solutions	45	.87	.78	.84

[a] n is the number of participants who provided data on this topic

The results for the retention of recommendations, problems and solutions are summarized in Table 3.11. Note that the number of recalls is rather low, especially for solutions (where many students indicated that they had no idea whatsoever) and the elaborations. The data was analyzed to see whether there were differences between the sessions one to four (that took place seven weeks before the post-session) and sessions five and six (that were five weeks before the post-session). The groups with the seven week interval indeed showed less retention of the material, but no such effect is visible for the recommendations. All differences have significance levels of .30 or higher, thus there is no indication that the additional two weeks to the post-session have biased the results for participants of the sessions 1 to 4.

Table 3.11
Counts of the Recommendations, Problems and Solutions Stated and their Respective Elaborations

	Stated	Elaborations	Total
	Recommendations (n = 61)[a]		
M	2.57	.89	3.46
SD	1.87	1.25	2.60
	Problems (n = 66)[a]		
M	1.92	.50	2.42
SD	1.19	.77	1.60
	Solutions (n = 45)[a]		
M	1.62	.47	2.09
SD	.91	.79	1.28

[a] n is the number of participants who provided data on this topic

Considering the drop in responses we focused further analysis on the retention of recommendations and problems and worked further with a total score for number and elaboration. In Table 3.12 the results of the retention of recommendations and problems and their respective elaborations are reported for the Collaboration and Detail conditions.

Table 3.12
Counts of the Total (Number plus Elaboration) for Recommendations and Problems

	Collaboration		Detail	
	Separate	Shared	Low	High
	Recommendations			
M	3.36	3.79	3.33	3.58
	(n = 47)	(n = 14)	(n = 30)	(n = 31)
	Problems			
M	2.38	2.56	2.90	2.05
	(n = 50)	(n = 16)	(n = 29)	(n = 37)

The results suggest that there may be a condition effect. Whereas we found that in the Shared Collaboration Mode fewer objects and relations were entered into the diagrams, the retention data reported here suggest that this did not lead to less processing of the material or the joint recommendations. On the contrary, we find that participants who were in the shared collaboration conditions recall more recommendations, problems and their respective elaborations. The results for the Detail conditions are mixed: The retention of recommendations for High Detail is higher than for Low Detail. The reverse however, occurs for retention of problems. The mean retention results are broken down per condition in Table 3.13.

Table 3.13
Recalled Recommendations (Number and Elaborations) by Detail and Collaboration Mode

	Detail Condition			
	Low		High	
	Separate	Shared	Separate	Shared
	Recommendations			
M	2.96	5.20	3.82	3.00
n	25	5	22	9
	Problems			
M	2.61	4.00	2.19	1.70
n	23	6	27	10

Multivariate analysis of variance indicated significant effects for Detail (Λ = .876, $F_{(2,52)}$ = 3.67, p = .03) as well as for the interaction between Detail and Collaboration mode (Λ = .891, $F_{(2,52)}$ = 3.19, p = .05). Subsequent univariate tests show that the significant effects are due to the effects on recall of the problems. Both the main effect of Detail ($F_{(1,53)}$ = 7.33, p =.01) and the interaction between Detail and Collaboration Mode ($F_{(1,53)}$ = 5.938, p = .02) are significant, that is (1) Low Detail recall of problems is significantly higher than High Detail recall and (2) the interaction between Detail and Collaboration Mode as shown in Figure 3.2 is the result of the different trends in the Shared condition. Low Detail recall shows an increase from Separate to Shared Collaboration, whereas High Detail recall decreases.

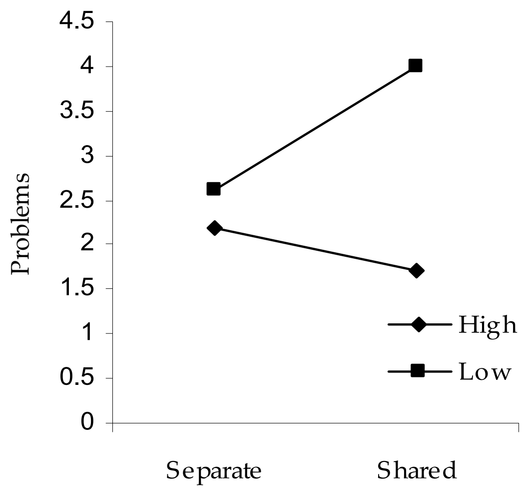

Figure 3.2. Recalled problems by collaboration mode and detail condition.

In summary: We found a main effect of the detail condition on the recall of the problems. More problems and/or their elaborations were recalled in the Low Detail condition, in particular when participants worked in Shared Collaboration Mode (the interaction effect). Again, the results contradict the predictions: higher specificity and a richer ontology did not lead to a better retention and recall of recommendations and problems. Our data indicate that the opposite effect is closer to the truth: Low Detail is associated with more recall, especially for Shared Collaboration conditions.

Retention of material

At the post-session we asked the participants three factual questions on particular parts of the background material that they had used in the experimental session. For example "Scheepstra did research on social integration of children with Down syndrome in regular education. What were her main conclusions?" or "What are the objections of the Flemish Board of Education against effect evaluations and what type of research are they proposing as an alternative?" The hand-written answers to these questions were typed and scored by the coder and the author using a detailed scoring protocol that assigned points for particular topics in the answers. In general, the participants recalled very few items of the material. This is reflected in the number of data available (declining from 40 to 12 answers) which also influenced the agreement between the coders: correlations for the scores on the three questions were .815 (n = 40), .411 (n = 34) and -.127 (n = 12). The summary statistics for the scores on retention of background material are given in Table 3.14. The numbers of available data and the low means indicate that participants had a very limited retention of the background material (Note: Students were not instructed at the time of the experimental sessions to try to remember the material).

Table 3.14
Summary Statistics for Retention of Background Material

	Question 1 (n = 40)	Question 2 (n = 34)	Question 3 (n = 12)
M	.50	.69	.08
SD	.72	.58	.29

We analyzed the effects of condition on retention of material and got some puzzling results: Whereas on question 1 the Low Details scored significantly lower than the High Details, no difference whatsoever was found on question 2. The type of question was the same and the content of the questions does not give a hint. However, the retention scores are so low in general that we should not put too much emphasis on their meaning even if there is, statistically speaking, a significant effect.

Discussion

There is one clear picture emerging from the data collected in this experiment: the ontology and specificity, at least as operationalized here in the Detail condition, had none of the predicted effects. We found no effect on the perceived difficulty, no effect on the recall of the material, and no effect on the evaluation of the time available for the task. Further, no effect could be demonstrated on the richness and connectivity of the Belvédère® diagrams produced by the participants (Low Detail was rich and exhibited more connectivity). Finally, no effect of the experimental condition could be demonstrated on the recommendations and problems (and their elaborations) compiled during the experimental session or their recall several weeks later. Although the experiment was hampered by technical problems, these results come out very consistently in all parts of the data. One of the few statistically significant effects found pointed in an opposite direction: Low Detail recall of problems was higher than High Detail recall and even more so in a Shared Collaboration Mode.

Most teams worked in 'separate' mode with team members sitting at their own computer running a Belvédère® client. As stated previously, for two sessions of the experiment this condition could not be realized and here Shared Collaboration Mode was introduced, an 'accidental' condition that however mimics the conditions under which most research with Belvédère® is done. We found that students working in Separate mode in general entered more objects and relations in the diagrams than those who worked together using one computer (Shared mode). The results corroborate earlier results found with Belvédère® (Suthers, 1999): when students work together using the same computer they tend to use the diagram to *consolidate* the results of their discussion, rather than to bring forward their individual opinions. This may result in fewer objects being added to the diagrams.

Subjects working in this Shared Mode have another advantage in that they have more powerful coordination mechanisms (i.e., plain ordinary speech and gestures), whereas in the Separate Condition students have to use a chat window to do the coordination. Veerman notes that students using Belvédère® needed to spend a high amount of effort to co-ordinate their communication and they "(…) spent half of their messages to technical issues and planning aspects, mainly considering the co-ordination of actions in order to construct a diagram" (Veerman, 2000, p. 121). This observation suggests that the communication service that Belvédère® offers (a chat-box) is at odds with the communication demands of the task. In chapter 2 we identified a demand for issue-based, conversational type of communication for the problem orientation state. For the problem solving state a demand for a topic-based type of communication was formulated. We can see that Belvédère® offers a mixture of conversational communication in the chat box and topic-based communication in the diagram. However, there are no facilities to create a link between objects in the diagram and the chat box. Moreover, the diagram window only contains a representation of an argument, one cannot use it to discuss planning or co-ordinate actions as would an agenda with actions to work on the diagram.

We, unfortunately, do not have the data on the chats or the dialogues that allow us to determine whether the contributions of objects and relations to the dialogue differentiates between the conditions (for example because students working separately generate more ideas), or whether the coordination processes make the difference (the same amount of ideas is generated but more of them are discarded in the coordination dialogue and never enter the diagrams).

The experiment failed to produce evidence for a number of predicted effects of specificity and ontology. Most of the data in fact pointed to directions contrary to the predictions. A number of alternative explanations for the lack of effect of specificity and ontology can be offered. In the first place we may consider the idea that the complexity of the High Detail condition requires so much additional coordination that students enter (or keep) fewer objects and relations in the diagrams. Even without studying the actual coordination processes of the learners, this is not a very likely explanation, because even in the *shared* condition where students can coordinate more easily we find that Low Detail is accompanied by higher richness and connectivity.

A more basic explanation is that students in the High Detail condition simply require more time for decision making on the categorical choices. This however is at odds with the reports of the students themselves: those in the High Detail condition were more positive about the available time than those working in the Low Detail condition.

A second type of explanation to consider is that our *definition* (i.e., our representational notation) of the objects in Belvédère® did not focus students or did not help them in making the categorical choices they needed to make to complete the task. Remember that in the Low Detail condition participants were instructed to only use *hypothesis* - to represent assumptions - and *data* to represent facts. In the High Detail condition objects were added that differentiated between analysis and the formulation of a solution or at least drew attention to the distinction: the meaning of the hypothesis symbol was illustrated with respect to implementation problems and the principle object was defined as 'solution to problems'. It seems that in the definition of the objects a number of relevant distinctions are being made but not sufficiently to support the students in formulating recommendations. In the instructions to the participants we stipulated that they were to present recommendations on the implementation of inclusive education, not their vision or opinion on inclusive education. We instructed participants to consider problems and dissenting views in their recommendations, but as we saw in the analyses, students formulated only a few problems. They concentrated on mapping the background material and often formulated - as their solution - a position ('we are in favor of inclusive education') and an example implementation ('as done in…'). This suggests that more is needed than changing the representational notation alone – in particular process support may be necessary, cf Kanselaar et al. (2003) on computer-based environments to support collaborative argumentative writing.

Conclusions

The experiment presented here has some 'one-shot' features: we used an existing system with its symbols, although we changed some of the meanings, to support solving a different type of problem than the system was designed for. In the High Detail condition we added some interpretations to the symbols to make them more usable for the 'how'-aspects of problem solving. One may argue that the results reported here were, certainly with the wisdom of hindsight, predictable: one cannot assume that minimal changes to a system for scientific inquiry will make it usable in other contexts and especially for other purposes.

The results reported and an informal inspection of the content of the recommendations give no reason to assume that adding a more specific and somewhat richer set of objects to an environment such as Belvédère® will ultimately lead to better solutions. Both the content of the representational notation as well as the process in which its use is embedded need scrutiny. Considering the type of problems that we are interested in, the representational notation and the process are to be derived from areas like social science problem solving or evaluation, rather than that of scientific inquiry. Such a representational notation was presented in the previous chapter, but it is more complex than the one used in the current study. There are reasons however that lead us to believe that it is not necessary to directly abandon the use of a more complex scheme. It might, in a different setting, still yield good results. A more dedicated scheme would signal to the students what sort of topics they need to address. This sort of process support seems desirable, if students, as was the case with the second year students of this study, had little prior knowledge of the specifics of the problem. In the next chapter we present a first, modest step to research the usability of a richer representational notation among beginning experts.

4 Towards a representational notation for social science problem solving[7]

One of the conclusions of the previous chapter was that students needed more than an add-on notation for tackling 'social science' problems in a CSCL-environment. We also concluded that they encountered problems because they had insufficient knowledge of the problem area. It is clear that direct translation of an external representation with an inquiry ontology to a different one (social science problems) is not the answer. At this point the research took a different route, namely the development and testing of a notation instead of an external representation tool for asynchronous CSCL.

Here we report on a first attempt to validate a representational notation that may support students who collaboratively solve 'social science' problems or analyze proposed solutions to these sorts of problems. The representational notation was introduced in chapter 2 and here we are concerned with its usability. As a first step we are interested to know whether students, without being instructed in the notation, already use concepts that are part of the notation. The idea is that if the concepts are already used by the participants, then the representational notation will be easier to learn and use. We do not expect students to *use* the *exact* terminology of the representational notation: they may, for instance, refer to 'actions' rather than to 'interventions'. The essential question is whether the *concepts* of the representational notation can be used to express the content of the student dialogue and whether these concepts help us understand their problem solving process. If that is the case, it supports the idea that the representational notation contains concepts and distinctions that the students are able to handle and that these are important concepts to help students solve these problems. The next step then is to incorporate the representational notation in an environment where students collaboratively solve problems. Figure 4.1 presents the notation in all details.

[7] We like to express our gratitude to prof. dr. Tom Duffy, Indiana University, for making available the data on which we report here. The development and testing of the coding scheme were done in cooperation with dr. Chad Carr. A shortened form of this chapter was presented at CSCL 2003 (Kirschner, Van Bruggen, & Duffy, 2003).

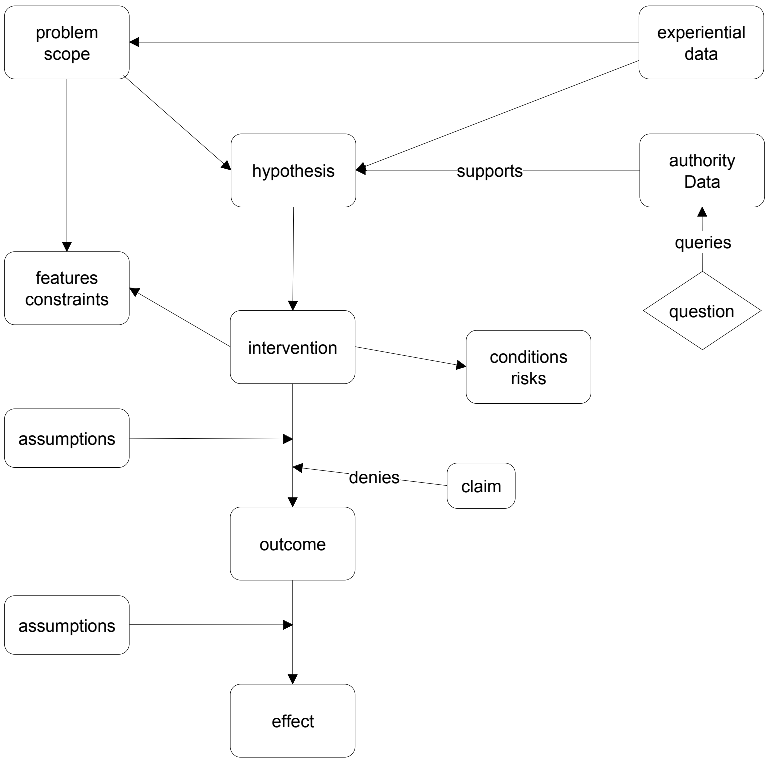

Figure 4.1 A Decision Representation Language (DRL). representation scheme for analyzing complex problem solving.

Design and methods

Method

To initially determine the usability and usefulness of the representation notation we discussed in chapter 2, we used it to analyze a small set of discussion protocols of students trying to find a solution to the continued increase in the number of drop-outs from school. The participants were 9 American university students with different educational backgrounds (debating, philosophy, education; each group consisting of three participants with the same background). The groups were selected because of the specific expertise that they were supposed to bring to the task. The Educators brought content knowledge of the educational system and educational policies. The two other groups brought process knowledge to the tasks. The Philosophers supposedly brought expertise in the areas of systematic analysis, valid reasoning and argumentation. The Debaters were supposed to bring process expertise in systematically discussing problems and arguing for specific solutions.

The participants were confronted with the following problem: The annual meeting of the Board of State Governors has asked a team of experts, you, to prepare recommendations to reduce school drop-out and to present those recommendations to the Board within a month. Participants were further instructed that they would have one other meeting shortly before the presentation to the Board of Governors, and that they could compile a list of research questions to be answered before that meeting. Their goal was to prepare a 10 minute presentation for the governors.

We prepared an analysis scheme based on the external representation described above to code the content of the typed written protocols of the first 90-minute sessions (see the Appendix). The scheme used four main categories: (1) problem analysis and problem scope; (2) features and constraints for the solution; (3) interventions (including expected outcomes and assumptions) and (4) the planning of the presentation. Each category had several subcategories. In the coding scheme all statements relating to a topic, including critical remarks or questions are scored. The coding scheme with examples for the categories and subcategories are in the Appendix.

The typed-written protocols were segmented on the basis of the *speech turns* taken in the dialogue. This is a somewhat 'defensive' segmentation strategy and certainly one that could lead to loss of data if subjects expressed different topics within one turn. It has the advantage of being objective, whereas alternatives would have implied that we would either have to rely on the segmentation put in by the typists (segment on punctuation) or use a fine-grained segmentation based on clauses, many of them, likely, containing elaborations of the same content.

Since a turn may address different topics, coders were instructed to code for the main idea expressed in the turn. In addition, we asked them to indicate where they would have preferred to segment a turn to make it easier to code. Their additional segmentation (less than 10% of the turns) was, however, not used in the analysis reported here.

Each of the three protocols was coded by two of the three trained coders that were available. The coders who were unaware of the backgrounds of the participants each coded two protocols. The coding was based on the typed protocols only (no audio or video data were available). Coders were instructed that, only in cases where the protocols contained explicit notes on intonation, they could use these notes to derive the meaning of interspersed affirmatives, negations, et cetera. The turns themselves were all coded as 'other'. We made no attempts to have the coders resolve disagreement. The results reported here, with the exception of reliabilities, are based on one coded protocol for each of the groups.

Results

Quantitative results

Coders classified the segments in one of four main categories (Analysis, Constraints, Intervention, Presentation). The category *Analysis* was used for statements relating to the scope of the problem, the causes of the problem, including statements about these causes (hypotheses) and evidence (data) brought forward in support of these hypotheses. In this category we also included statements about principles and the research questions that could be formulated. The main category *Constraints* was used to code for statements about desirable or necessary features as well as preconditions and risks. The main category *Intervention* was used for statements dealing with interventions proposed, their expected short-term and long-term effects and statements about how the interventions would lead to these effects. The *Presentation* category was used for statements relating to the preparation of the presentation that the subjects, according to their instruction, had to deliver. Further details and examples are given in the Annex.

Each main category had two to six minor categories. Agreements between the two coders on the main categories was 72%, 77% and 79% for the Philosophers (P), Educators (E) and Debaters (D) protocols. Kappa's for the protocols were .59 (P), .66 (E) and .69 (D) respectively, which is sufficient to further explore the data. Agreements for the minor categories range were 66% (P), 69% (E) and 72% (D). The Kappa's for the total of minor categories were 0.53 (P), 0.57 (E) and 0.62 (D); per category the Kappa's differ greatly.

Table 4.1 presents the number of segments coded overall and the subset of those segments in one of the four main categories. As the table shows, the coders were able to classify over 50% of the segments into one of the four main categories. The number of segments, the total number of words and the mean number of words per segments in the protocol, reported in Table 4.2 indicate that the Debaters were considerably more verbose than the Philosophers and Educators, using roughly twice as many words across the segments. However, as the stereotype would have it, the Philosophers tended to produce much longer turns – at least 50% longer than the Educators or Debaters.

Table 4.1
The Proportion of Segments Classified into each of the Main Topic Categories or as 'Other' for the Philosophers (P), Educators (P) and Debaters(D)

Category[a]	P	E	D
Analysis	12.4	19.5	22.8
Constraints	14.8	3.3	1.4
Interventions	21.6	15.0	25.0
Presentation	5.6	11.9	2.8
Other	45.6	50.2	48.0
Segments[b]	250	452	921

[a] Category scores are in percentages
[b] Segments are absolute number of segments in the protocol

Table 4.2
The Number of Segments, Total Number of Words and Words per Segment for the Philosophers (P), Educators (E), and Debaters (D)

	P	E	D
Segments	250	452	921
Words	7548	8479	15098
Words/segment	30.19	18.76	16.38

Table 4.1 presents the proportion of segments that fall into each of the four main categories as well as those classified as 'other'. Each of the segments classified in one of the main categories was classified into between two and six subcategories as reflected in Table 4.3. The percentages are in terms of the total for valid categories, that is the total excluding the turns that were scored 'Other'.

Table 4.3
The Percentage of Segments Classified into each of the Topic Subcategories for Philosophers (P), Educators (E), and Debaters (D)

	P	E	D
		Analysis	
Data Authority	2.2	.9	.6
Data Experiential	4.4	23.2	18.6
Hypothesis	2.9	10.3	14.2
Principle	.0	.2	.0
Problem Scope	9.6	3.1	10.2
Research Question	3.7	1.8	.0
		Constraints and Features	
Features	16.9	4.0	1.7
Risks	10.3	3.1	1.0
		Interventions	
Assumption	4.4	1.3	7.9
Effects	.0	.0	1.9
Intervention	30.9	25.9	32.4
Outcome	4.4	2.2	6.1
		Presentation	
Content	4.4	2.2	2.1
Goals	.0	4.5	.6
Plan	5.9	17.4	2.7

Inspection of the minor categories shows that some type of content was rarely mentioned in one or more of the protocols. For instance few statements were coded as *Data Authority*. This might be attributed to our strict definition which only accepted stated sources as authorities, yet even in the protocol of the Educators – who had a beginning expert on the subject of drop-out on their team – there is little explicit mention of data based on authorities. Although this applies to the Philosophers in particular, none of the groups seemed very much concerned with stating or gathering (viz. the frequencies of *Research Question*) data based on authorities. The category *Principle* was posing problems to our coders. It was meant as a warrant to link the relation between observables (data) and drop-out. Coders found it hard to identify warrants and, of course, warrants are often left implicit in dialogues (Toulmin, 1958).

The subcategories in *Constraints and Features* show remarkable differences between the groups that we will discuss in more detail below. The categories under the heading *Interventions* deal with the program logic of the proposed interventions: what actions will be undertaken, what are the expected short-term effects and how will these actions lead to the long-term effects? What we see in the data for this category is that all groups expend considerable effort in discussing interventions (which is of course what we asked for) and very little in discussing the final effects (which were given: reduce drop-out). It is interesting to see the differences between the groups in the effort spent on relating interventions to the ultimate effect. Here again, we may hope that some guidance can be offered to help students realize that they must argue why an intervention is likely to produce the intended results.

Table 4.1 and Table 4.3 suggest that the groups followed different approaches with the Educators and Debaters concentrating more on *Analysis* and the Philosophers concentrating on the *Constraints* of the solutions. Proportional data, however, are not the best metric for making comparisons. We have therefore rank ordered the data in Table 4.3 from most to least frequently occurring segments for each of the three groups of problem solvers. Those rankings are shown in Table 4.4. For ease of comparison, we have prepared a separate table (see Table 4.5) with the 'top 5' rankings in each group.

Table 4.4
The Rank of each Subtopic Category in Terms of Frequency of Occurrence in the Protocols of the Philosophers (P), Educators (E), and Debaters (D)

	P	E	D
		Analysis	
Data Authority	12	13	13
Data Experiential	7.5	2	2
Hypothesis	11	4	3
Principle	14	15	15
Problem Scope	4	7.5	4
Research Question	10	11	14
		Constraints and Features	
Features	2	6	10
Risks	3	7.5	11
		Interventions	
Assumption	7.5	12	5
Effects	14	14	9
Intervention	1	1	1
Outcome	7.5	9.5	6
		Presentation	
Content	7.5	9.5	8
Goals	14	5	12
Presentation plan	5	3	7

Table 4.5
The Five most Frequently Occurring Subtopic Categories (those Unique to a Group in Bold) and Subtopic Categories Missing in the Top Five for a Group but Present in the Top Five for the Other Two Groups.

rank	Philosophers	Educators	Debaters
1	Intervention	Intervention	Intervention
2	Features	Data	Data
3	**Risks**	Presentation Plan	Hypothesis
4	Problem Scope	Hypothesis	Problem Scope
5	Presentation Plan	Features	**Assumptions**
missing	Hypothesis (11)[a]	Problem Scope (7.5)	Features (10)
	Data (7.5)		Presentation Plan (7)

[a] Numbers in parentheses indicate rank order scores for missing

The top-5 rankings for the groups represent 73.5% (P), 81.3% (E) and 83.3% (D) of the non-trivial (= *Other*) turns in the protocols. In Table 4.5 we have typified the groups by indicating their unique top ranking (in bold). The Philosophers have a unique high ranking for *Risks* and the Debaters for *Assumptions*. A group is not only typified by unique characteristics it possesses, but also by characteristics that it lacks. We have therefore added the unique *missing* categories, i.e., the other two have it, and indicated these missing categories along with their ranking (in parentheses).

Table 4.5 shows us that all groups focus on interventions, which is what we could expect considering the instructions given to the participants. The Educators and the Debaters share a focus on *Hypotheses* and *Experiential Data*, that is the causes of drop-out and data to support or refute these hypotheses. In contrast *Hypothesis* has rank 11 for the Philosophers. Their focus is on *Constraints*, *Features* and *Risks* of the solutions. Both Debaters and Philosophers focus on *Problem Scope* (rank 4), which was at rank 7.5 for the Educators. The Debaters spend less time on *Presentation Plan* (rank 7) than the other groups.

The picture that emerges from these rankings and the missing categories is that of different problem approaches. The Philosophers approach the problem, not by analyzing its causes, but by concentrating on the scope of the problem and defining desirable features of the solutions as seen in their rankings for features and problem scope. Put slightly differently: their work is focused on constraints and criteria by which proposed solutions can be evaluated. Both Educators and Debaters focus on causes, but the Educators don't seem to address problem scope, but rather focus on the presentation (as seen in their high ranking for *Presentation Plan*). Debaters share with the Educators the focus on analysis and interventions, yet they seem more focused on establishing the scope of the problem and to discuss the outcomes of the intervention. This is seen in their high rankings for *Effects* and *Outcomes*. Preparing a plan for the presentation – which directly relates to the instruction given – is not among their highest ranks.

At this point we collide with the limitations of the coding scheme. Whereas the frequencies and rank orders reported give an idea about the relative amount of dialogue spent on the different *types* of content categories and thereby allow a glimpse of different approaches taken by the three teams, they cannot give us an indication of the depth and breadth of the content covered in the dialogues. In the metric used, stating three different hypotheses, that is *breadth* in terms of content topics, results in the same score as repeating one single hypothesis three times. Note that this limitation of the coding scheme does not apply to the intended use of the representational notation in collaborative problem solving, since we may expect participants to describe the content of an object.

In order to cope with the shortcomings mentioned, a fourth coder was used to identify the *content topics* in the segments that were previously coded as hypothesis, interventions or constraints and features. These categories seemed most promising to reveal details of the approaches used. The coder was not involved in the original coding of the protocols and she was unaware of the backgrounds of the groups. We asked her to identify *all* the *content topics* within the segments so her coding could lead to several codes per segment.

Table 4.6 reports the *number* of content topics, that is the number of different hypotheses, proposals for interventions and constraints that the three groups mentioned and how often these content topics were mentioned in the protocol (numbers in parentheses). The number of content topics formulated is an indication of the *breadth* of the approach. For instance, the Philosophers formulated two different hypotheses, whereas the Debaters formulated no less than ten hypotheses. The number of segments in which a content topic occurs is an indication of the *depth* of its coverage. The Philosophers, for example, referred no more than four times to their two hypotheses. In contrast, the Debaters formulate ten hypotheses to which they refer in 90 segments.

Table 4.6
Number of Different Content Topics for each of the Main Categories and Number of Segments referring to them (Between Parentheses)

	Philosophers	Educators	Debaters
Hypotheses	2 (4)	9 (25)	10 (90)
Interventions	8 (42)	9 (54)	5 (124)
Constraints	11 (37)	8 (11)	11 (38)

Table 4.6 corroborates a number of observations made before. First, it shows very clearly that the Philosophers concentrate on different constraints for their solution rather than on the potential causes of drop-out. First, they formulate no more than two different hypotheses and address these in four segments. In contrast to this, they formulate no less than eleven different constraints and refer 37 times to them. In the second place the table gives us a first hint at the approach of the Debaters. The breadth of the Debaters and Educators is, by and large, comparable. The Debaters, however, address the content topics more often than the two other groups. As an example, consider Interventions: the Educators describe nine different interventions to which they refer 54 times. The Debaters mention five different types of interventions but refer to them a 124 times. If the number of different content topics together with the references to them make a good indicator of the importance of the topic category (hypothesis, constraint, intervention) then the Debaters seem to assign almost equal importance to hypotheses, constraints and interventions.

Unfortunately, the *aggregate* data in Table 4.6 conceal that some content topics get no follow-up in the discussion, whilst others are discussed over and over. Before making any statements about the actual breadth, depth and focus in the protocols, we have to check which content topics are really discussed. The following figures relate the number of content topics with the references to them, they plot the number of content topics (y-axis) against the minimum number of references (x-axis) made to them in the protocols.

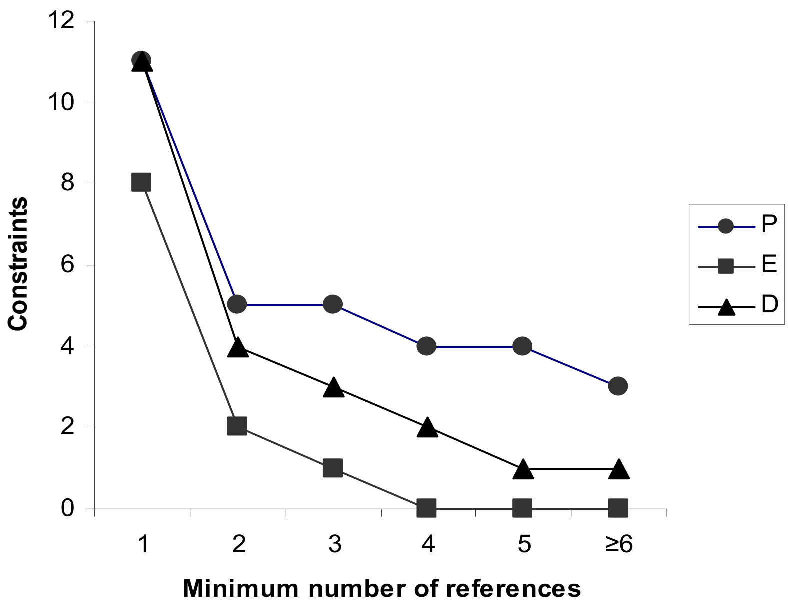

Figure 4.1. Breadth of coverage for Constraints: the number of constraints in relation to the minimal number of references made to them in the text for Philosophers, Educators and Debaters.

Figure 4.1 shows that the Debaters formulated eleven different constraints (eleven constraints were mentioned at least once), but only four of these are mentioned twice or more. The Educators mentioned eight different constraints but focused on two. All groups show the same type of 'decay function': a fast drop after the first occurrence and a steady decrease after that. Note that the Philosophers formulated the same number of constraints as the Debaters, but they kept more constraints in focus for a longer time.

Figure 4.2 on *Hypotheses* and Figure 4.3 on *Interventions* are the most interesting, because they clearly illustrate the different approaches of the groups. From Figure 4.2 the approach of the Philosophers seems clear: they pay little attention to the causes of drop-out: two different causes are mentioned, one of them three times. Compare this to the approaches of the Debaters and the Educators. They formulate almost the same number of hypotheses (10 vs. 9), but there is only *one* hypothesis that the Educators refer to five times or more. In fact, this single hypothesis is mentioned nine times. The reason why we only plotted the data that demonstrate the drop in focus will become clear when we consider the Debaters' data. As the figure shows there is hardly a drop in the number of hypothesis that the Debaters keep in focus. In fact, there is one hypothesis that is mentioned nine times and two other hypotheses are even mentioned 22 and 23 times. We would of course expect that the Debaters, whose protocol is much longer than that of the Educators, produce more references to the topics. It is the 'decay pattern' that is so different: the Debaters seem to keep a very broad focus on the causes, rather than concentrating on a small set of causes.

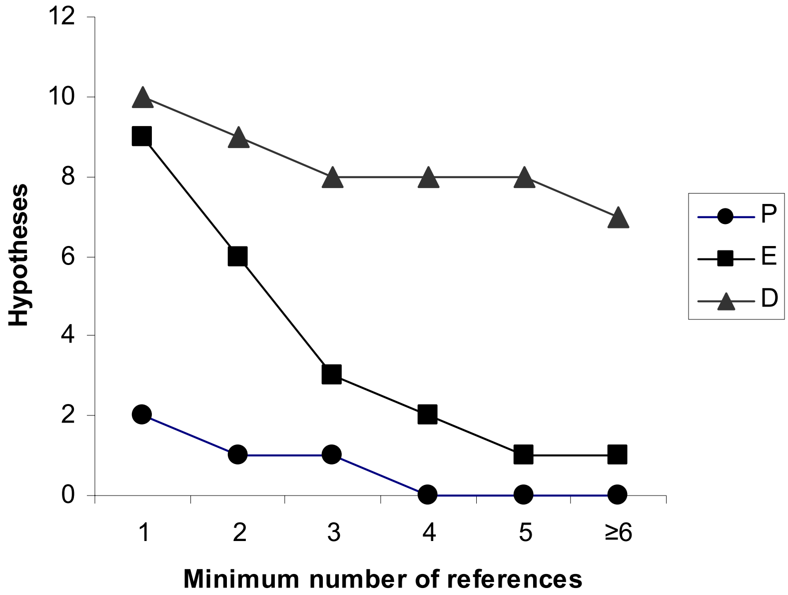

Figure 4.2 Breadth of coverage for Hypotheses: the number of hypotheses in relation to the number of references made to them in the text for Philosophers, Educators and Debaters.

The tendency to maintain a broad focus is also demonstrated in the next figure that has the content topics on *Interventions*. All groups concentrated on *Interventions* reflected in the number of references to the concept topics. The Philosophers and the Educators reduce their initial focus although less drastically than they did on the other topics. The Debaters again keep a very broad focus and keep returning to a larger number of topics. We cannot even plot the turns for the remaining four Interventions (they were referred to 13, 26, 30 and 47 times).

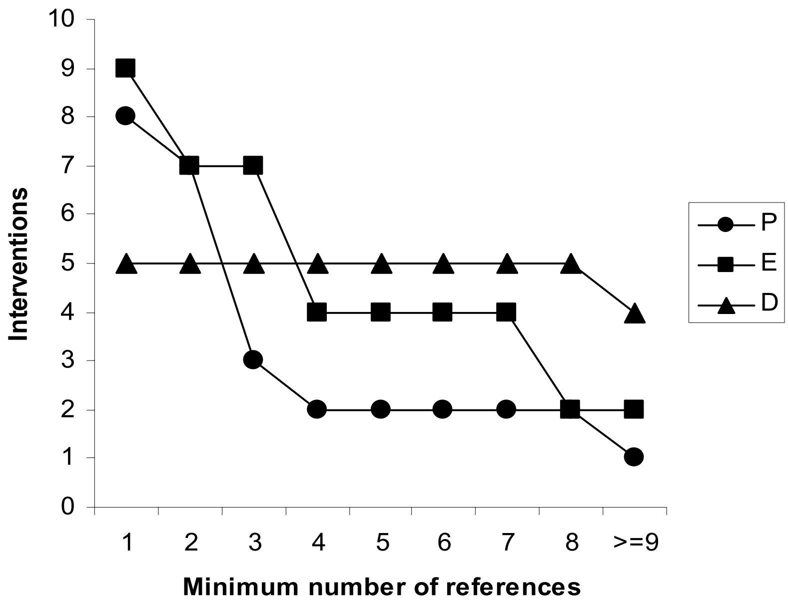

Figure 4.3. Breadth of coverage for Intervention: the number of interventions in relation to the number of references made to them in the text for Philosophers, Educators and Debaters.

To get a clearer picture of this supposed pattern, we plotted the occurrences of content topics in the protocols. In the following three figures we can view this pattern. The x-axis of figures 4.4 through 4.6 represents the Segment Number (0 is the beginning of the session, maximum is the end of the session). The Y-axis represents the topics discussed. The final letter of the Topic signals whether the topic was a Constraint (C), a Hypothesis (H) or an Intervention (I). The crosses and vertical strokes on the line of a topic represent the segments in which the topic was discussed. The horizontal strokes are added to increase readability. In Figure 4.4 qualC is the Constraint that the Quality of education may not be impinged upon. This was the second constraint discussed (after retrC) and we can see that it was discussed, left for a while (for about 30 segments) while other constraints were discussed and then returned to for a number of segments. In Figure 4.4 we see that the Philosophers formulate several constraints, most of them in the early part of the protocol. As we saw above several of these constraints get no follow-up, but a few are referred to more often. An example of the latter is their constraint that no solution shall compromise the quality of education (content topic qualC in the graph). A similar pattern can be noted in the content topics covering Interventions. The Philosophers have one type of Intervention that is brought up throughout the whole protocol. Others are more concentrated in smaller series of segments.

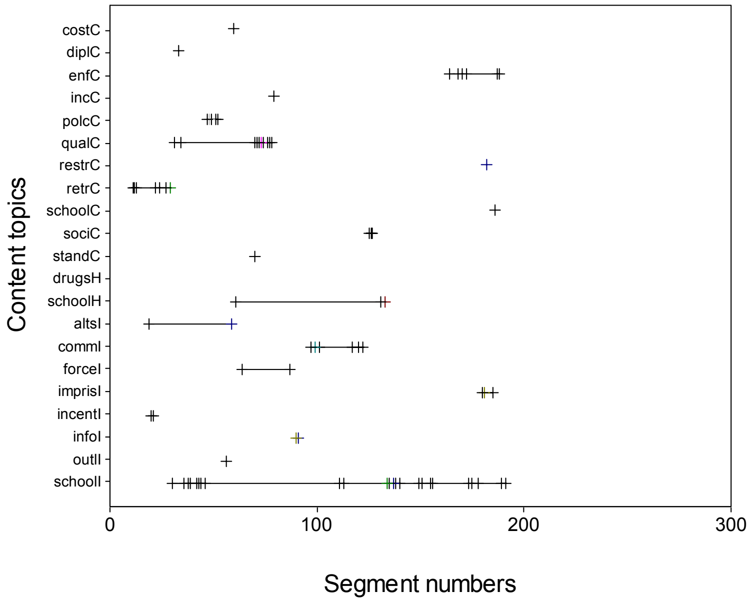

Figure 4.4. Occurrence of Constraints (the names on the Y-axis ending with C), Hypotheses (names ending with H) and Interventions (names ending with I) in the protocol of the Philosophers.

Figure 4.5 demonstrates the focused approach of the Educators (we omit Constraints from the graph, because the Educators and Debaters formulated only a few). In the first part of the protocol several hypotheses are formulated of which only two are kept in focus. From here several interventions are mentioned, but the Educators focus on three interventions until, in the middle of their protocol, a number of other hypotheses and interventions appear without any visible follow-up.

Figure 4.6 demonstrates the broad focus, or rather the lack of focus, of the Debaters. The Debaters, as inspection of the protocols teaches us, used the results of their initial brainstorm – a list of potential causes – as a guide for their further discussion. They worked down that list: systematically and in great detail, discussing implementation issues without checking the effectiveness of the interventions they discussed. If there is lack of focus in their protocol, there is method to it.

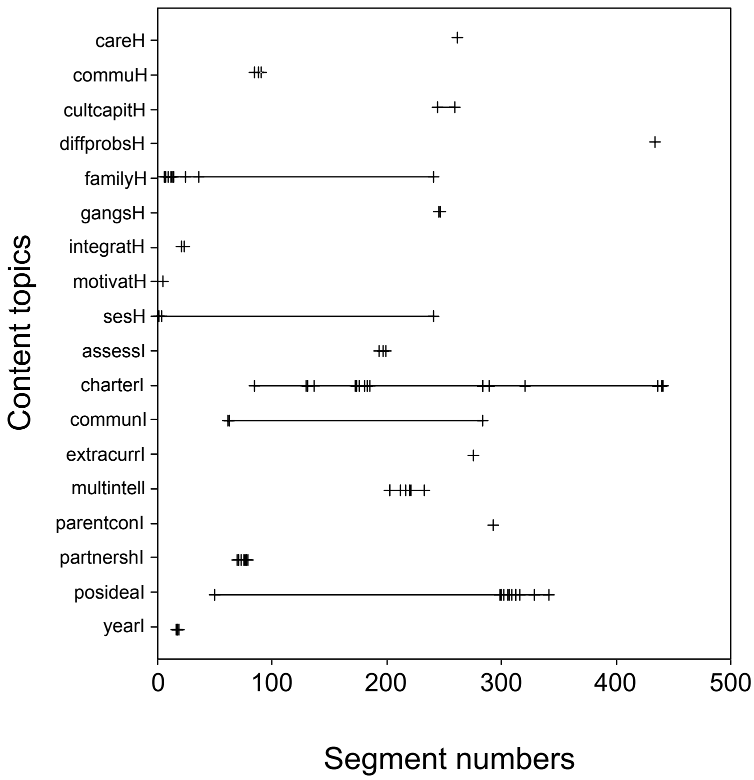

Figure 4.5. Occurrence of Hypotheses (the names on the Y-axis ending with H) and Interventions (names ending with I) in the protocol of the Educators.

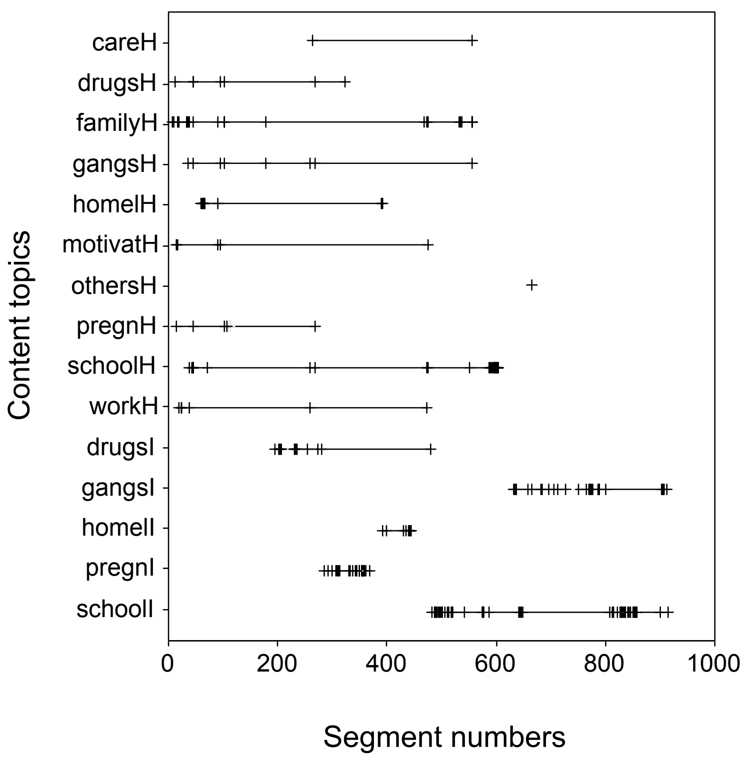

Figure 4.6. Occurrence of Hypotheses (the names on the Y-axis ending with H) and Interventions (names ending with I) in the protocol of the Debaters.

We started with the question whether students use the concepts in the representational notation when they engage in collaborative problem solving. We expected that a coding scheme based on the notation would account for a substantial part of student dialogues and would help us understand the problem solving process of the students. We found that the scheme can account for about 50% of the turns in the dialogues. The inter-rater reliabilities are modest, but acceptable for the current purpose. Improvements for the purpose of understanding dialogues may be achieved if we include codes for process control and reconsider the codes that were rarely used. The analysis of the content topics indicates that the groups followed different problem solving strategies that corresponded to a combination of concentration on *particular* content (solution features, or causes) and the breadth and depth of focus: the number of content topics and the number of references to them. To check these suppositions we carried out a qualitative analysis of the protocols that is reported in the next section.

Qualitative results

A qualitative inspection of the protocols corroborates the inferences drawn from the quantitative analysis of segments. The *Philosophers* discussion starts with delineating the problem: they discuss the meaning of drop-out and how high the retention rate should be [segments 3-18]. This leads to a first potential recommendation: expand the number of alternative high schools other than college prep schools. In segment 19 and following the Philosophers develop general characteristics for the intervention (incentive / punitive) [segment 20] and features, leading to what they call a meta-recommendation: no recommendation should "sacrifice the quality of education or the standards that someone has to meet in order to graduate" [segment 31]. This is the important constraint that emerged from the quantitative analysis. Their next step is to define targets for intervention: schools, parents and "governmental, societal, environmental" actions [segment 36]. They elaborate on this by giving examples of potential actions, noting that these will often be socially and politically difficult to realize [segments 37-56].

What the Philosophers are doing can be described as structuring the solution space in terms of constraints and features, target groups and intervention methods. Beyond their initial scoping statements there is no sign of an analysis of the causes of the problem or other problem structuring activities in the beginning of the protocol. In segment 57 they formulate a research question on drop-out figures related to geographical area and background data, such as race, gender, class, and in segment 102 another research question is stated as to "what programs and things worked and what are the best sort of incentives (…)". It is only at segment 130 that they start considering causes. Before that they have considered in some detail incentive approaches, concluding that there will be a huge pressure on the teachers to award the incentives, and punitive interventions, concluding that a similar pressure will invalidate this approach as well.

The approach of the *Debaters* is completely different. They start with a brainstorm on a number of possible causes (drugs, gangs, pregnancy, jobs, homelessness, et cetera) adding some experiential data to support these claims (see Figure 4.5). Their subsequent approach is best summarized in segment 74: "… we have to like, *find solutions to all these problems* and present it in 15 minutes anyway". After some debate they agree on including at least a recommendation dealing with the curriculum and then proceed to systematically work down the list of causes. For each cause on the list they try to define an intervention that will eliminate the cause or reduce its effects. The results on analysis and interventions that we previously presented are produced by them enumerating (and repeating) the list of causes, and discussing the details of the implementation of the interventions for each of the separate causes.

In their approach, the Debaters try to formulate school-based interventions for problems that they perceive as being more social in nature, and this leads to a sort of breakdown later on in the protocol (see discussion below). Interestingly enough, they often fail to formulate a particular recommendation (probably the reason why Features only reaches rank 8 in the coded protocol). For instance, the last segments devoted to the drugs as a cause of drop-out [segments 149-280] bring forward underlying causes, but without any noticeable influence on the overall approach. Similar discussions appear at the end of their discussion on how to reduce teenage pregnancy (they only consider distribution of free condoms through the school) [segments 282-385] and at the end of their discussion on homelessness [segments 386-435]. This leads them to reconsider how much they can actually achieve – "(…) we don't have social programs (…) or solutions for these social problems that's causing drop-out" [segment 442]. The Debaters then decide to narrow their scope – maybe the recommendations should target only those learners who can be helped directly by making changes to the school [segment 529-533], possibly even writing off other learners [segment 590]. The Debaters in the following part of the protocol come up with several school related recommendations: teacher quality, improve resources, more extra-curricular activities and others.

It is interesting to note how both the Philosophers and the Debaters run into problems caused by their initial approach: the Philosophers don't analyze causes, but formulate constraints and intervention-strategies that turn out to be incompatible. The Debaters analyze the causes and try to define school-based interventions for each of them that turn out to be at variance with the social nature of the stated causes. Debaters also loose track of the larger goal of reducing drop-outs and begin to focus on how to implement (or whether to implement because of social considerations) the intervention. For example, the implementation of support for pregnant girls looses track of how big of a drop-out problem this is or even if the solution addresses the drop-out problem and focuses more on the practical and social issues in implementation.

The *Educators* obviously have more background knowledge to substantiate their analysis and recommendations and their protocol has lengthy passages in which (research) data are summarized and explained by one participant to the others. The analysis of the problem is very short, or even shallow. A number of hypotheses are formulated: low social economic status, families and communities, motivation, but already in segment 21 the core of their analysis and recommendation is formulated: community and parental involvement. Authority data are presented that relates involvement to reduction of drop-out. That participant, who is a beginning expert in the field, points out that parental involvement is a goal in a program most governors have already adopted [segments 32-34]. Even when explicitly asked for innovative approaches or approaches that have failed [segment 42] the dialogue turns into reaffirming the importance of community involvement, and after a number of short exchanges the core recommendation, stimulate foundation of charter schools, is introduced in segment 70. This is followed by a series of lengthy exchanges in which that participant gives all sorts of information on charter schools, while another participant tries to formulate the points that are appealing to the governors [segments 71-126].

In segment 131 the Educators start planning their presentation and anticipate reactions of the governors, including the issue of standardization [segments127-219]. This leads to a number of innovations they would like to include in their recommendations (alternative assessment, portfolios, multiple intelligences) although the relation with drop-out is not very clear. In segment 220 the issue of drop-out is raised again and a participant asks for other approaches and programs that could be included. It is here were the Educators briefly consider causes mentioned by the Debaters: family problems, low opportunity areas, crime, gangs et cetera. They make an important scope decision to not address external factors beyond their control [segment 258] and stick to their educational recommendations.

There is some sort of 'confirmation bias' in the Educators protocol: no alternative causes or interventions receive substantial attention and the participant who questions whether the recommendation would reduce drop-out is mostly ignored. The Educators seem to test their recommendation not on whether it will actually reduce drop-out but on whether it will fit stated policies of the Governors.

Discussion

The results presented here show resemblance to the analysis of social science problem-solving by James Voss and his colleagues (Voss, Greene, et al., 1983; Voss, Tyler, et al., 1983). Voss, Greene, et al. had experts, novices and beginning experts solve the problem of how to increase Soviet agricultural production. Experts described the causes of the problem in more abstract terms (infrastructure) and they stated more general abstract solutions to which lower level problems were subordinate. Novices decomposed the problem in low-level sub-problems to which solutions were proposed; an approach that the Debaters (who are novices in the area of expertise needed) seem to follow. Novices failed to evaluate solutions in terms of constraints and they did not specify sub-problems that could be encountered when proposed solutions were implemented. Failure to evaluate solutions against constraints is also found in the Debaters' protocol and the sub-problems that they identify emerge during their discussion on how to implement a recommendation. The Debaters discussion is guided by the potential causes that they mentioned during their initial brainstorm. They systematically work down this list, without checking for effects or constraints. The Philosophers, as did Voss et al's experts, stated constraints and general solutions and tested these solutions against the constraints (and found out they were incompatible). In contrast to Voss' experts the Philosophers and the Debaters did not relate their solutions to an abstract problem definition. The Philosophers did not analyze the causes of the problem to any depth and the Debaters in essence generated a list of *unrelated* causes and, working down this list, tried to formulate interventions for the individual causes.

Finally, the Educators, whom we may consider beginning experts, more or less structure the problem space in the expert way, that is they soon concentrate on two underlying factors that they consider crucial in school drop-out: community and parental involvement. The constraints they adopt are to the point and pragmatic: they test the solutions that they propose against policies that the governors have already adopted. Their solution is strictly educational – although this was not a stated constraint – and when they realize that other factors may be important as well, they explicitly decide to restrict the scope of their solutions to the educational arena.

There are some indications that the participants followed the problem solving states and the associated communication styles identified in chapter 2. Both Educators and Debaters start with a brainstorm on potential causes. The Educators hit on their solution within minutes however. The Debaters start with a rapid brainstorm on the causes of drop-out. They then prioritize this list and use it as an agenda of topics to be further explored. A substantial part of their protocol consists of defining the interventions that will eliminate the individual causes and finding ways of implementing these interventions. The Debaters' protocol demonstrates that a topical agenda (as was suggested for Belvédère® in the previous chapter) alone will not suffice as support and guidance. The Debaters had the topics to discuss but got entangled in the details of implementation. They obviously had no, or failed to apply, a stopping rule, such as the level of detail required for a recommendation to the Board of Governors. Moreover, they did not consider the relation between cause and drop-out and the effectiveness of their recommendations. The Debaters worked systematically, but as we analyze their activities from the perspective of our representational notation, we see that they address topics in isolation. In a similar way the Philosophers worked on the constraints that their interventions needed to satisfy, without analyzing the problem in any detail. All this demonstrates that having a topical agenda by itself will not suffice, we need guidance and support that stimulates students to work on the whole fabric of constraints, causes, data, interventions and outcomes to only mention the core objects.

Now all this was done as a first step towards the use of a representational notation to support collaborative problem solving: we have attempted to determine whether we could use the scheme to analyze problem solving protocols. Before we present our conclusions with respect to usability, we need to enumerate a number of limitations associated with the approach that we have taken.

First, we have concentrated on the content of the dialogues and tried to model it using the concepts of the representational notation. To be more specific: we used the objects, and not the relations, of the representational notation, and as a consequence of that approach we cannot model the structure of the argumentation of the students. Furthermore, we have analyzed the individual segments only and have not looked at larger structures such as episodes devoted to a particular topic. The way in which the student dialogue progresses through a number of topics and how the students maintain focus are not modeled in the coding scheme used. However, the analysis of the content topics, gives us an accurate estimate of the focus for particular types of content.

Second, the reliability of the coding schema is a concern. The moderate Kappa's point to potential conceptual ambiguities in the coding scheme and indeed the coders reported difficulties with certain distinctions. They felt that some categories were lacking such as a category to cater for dialogues in which actions to be taken are discussed. They also mentioned a need for a category, such as "Intervention Strategy" to cater for statements dealing with *types* of interventions rather than concrete, individual cases. These problems are not related to lower reliabilities however (the new categories would be coded as 'Other' and 'Intervention'). Coders, however mentioned that they found it hard to make a distinction between constraints and features. Now it may indeed be difficult to decide from the protocols alone whether a participant is considering something a constraint or a feature. The ultimate criterion is, however, not whether the coding of the protocols is reliable enough, but whether students can use the objects of a representational notation in a systematic way. For the current analyses, however, the Kappa's nevertheless remain a concern.

Third, there are several important aspects of the dialogue that we do not capture, for instance the means (speech acts) by which students express their ideas and arguments, how they coordinate their actions, and how they negotiate knowledge. We expect that the use of an external representation will eventually help students to maintain focus and coordinate their actions, but this is not modeled in the coding scheme that we used.

Finally, the empirical base for the different approaches identified is small. We may be able to add to it however. In December 2002, too late to analyze and report the results here, we received material from six parallel groups (two Philosophers, two Debaters and two Educator groups) who had worked on the same assignment.

Conclusions

The analysis presented here is the result of a first step towards the use of a representational notation to support collaborative problem solving. In this first step we are interested to know whether the concepts in our coding scheme correspond to those used by students when they solve problems. In the long run we want to find out whether the representational notation can support collaborative problem solving.

In our analysis we found that the coding scheme can account for an important part of the content of the dialogues. In other words, students use, at least, a number of the concepts that are behind the representational notation. That seems to make it more likely that students can master the objects and relations contained in the notation. However, the reliabilities that we observed force us to be more cautious in these conclusions.

We noticed that there were important differences in the problem solving strategies of the three groups. The representational notation must be robust enough to support these different approaches. First, we noticed that the coding scheme could cater for the contents of the three teams. Moreover, we could typify their problem solving approaches using the concepts contained in the coding scheme. The approaches of the Debaters and Philosophers can be described in terms of the particular objects they ignore and the subset of objects they focus on. The Debaters, for example, got bogged down into implementation details of a set of unrelated causes that was the result of their initial brainstorm. They ignored the relation of the interventions they considered with drop-out (effects) and they ignored constraints on the nature of their recommendations (what sort of recommendation do the governors expect). Similar observations can be made regarding the Philosophers. They as well concentrated on particular aspects of the problem, in particular *constraints*, while ignoring others, in particular the causes, almost completely. Although their instruction made it clear that they had to compile and present recommendations to the Board of Governors, this constraint did not lead to any discussion in these two groups.

Considering that the groups concentrated on particular contents, while ignoring other and that their recommendations were not very convincing, there is also an opportunity that in particular these groups may profit from more process support.

We do not assume that making available a representational notation alone, will be sufficient to provoke more systematic problem solving activities. As we described above, the support should stimulate to address the topics along with their relations. That is, a kind of guidance and support is needed that, for example, encourages the groups to engage in a more structural analysis of the problem as well as the solution. Students should be encouraged to formulate a more coherent and concise set of interventions that is checked against constraints (in this case, does it fit with known policies) and the intended effects.

In the next chapter we will further explore the issue of process support. There we will use multi-disciplinary teams that were trained in techniques to support the process.

5 Logical Framework Analysis as process support

In the previous chapter we took the first steps towards a representational notation for collaborative solving of ill-structured problems using non-novice students. We analyzed a number of protocols to determine whether students already used concepts that were included in the notation. Furthermore we checked whether absence of concepts indicated places were the problem solving process could be supported.

Analysis of the protocols showed that the students participating in the experiment demonstrated beginning-expert behavior in their areas of expertise: The Philosophers spent a considerable amount of their deliberations on defining the scope of the problem and refining the constraints of the solution. They more or less approached the problem as an ethical one and tried to delineate what solutions would be acceptable. The Debaters demonstrated their expertise by their methodical approach. They first brainstormed on possible causes, next they drew up a list of causes which they then prioritized and finally they worked down this list. They discussed, one by one, the causes and the potential solutions that might remedy the cause or alleviate the problem.

Despite their systematic approaches, both groups failed to produce acceptable solutions: The philosophers ended with a number of very weak recommendations and the debaters got entangled in details of implementation questions and forgot to check whether their recommendations had any credibility as to their effectiveness. The Educators, who had knowledge of the specific area, came up with a sensible solution within the education realm (foster creation of charter schools), but failed to look at non-educational aspects of the problem or their solution. We concluded that these results indicated that the non-education students would certainly profit from more process support in their problem solving activities and that this support had to emphasize the structural relations between the components of the analysis and solution. For example, support and guidance should stimulate that causes would be substantiated with data, interventions checked against stated constraints et cetera.

Process support can be offered in a variety of ways. Duffy et al. (1998) describe how they support Problem Based Learning by offering different modes of communication and types of representation (issue, topic) for the phases of Problem Based Learning. During an orientation phase, the problem-solvers may engage in conversation that is *issue-based*. Here they move freely from one issue to the other and make a first inventory of the problem and potential solutions. In a problem-solving phase the students are communicating *topic-based* so as to explore potential solutions more deeply. This is a phase where an asynchronous mode of work allows students to explore the topics more profoundly than the synchronous mode used in the exploration phase would allow. These observations were incorporated in the demands of the different problem solving states that we formulated in chapter 2.

Process support in the form of offering communication modes alone or a topic based agenda alone will not suffice. Consider the example of the Debaters of the previous chapter. They had an initial brainstorm, issue-based, that returned a list of possible causes. They then went to a topic-based approach, addressed one by one the causes on their list and formulated interventions for each of these causes to, finally end up with a set of unrelated interventions whose potential effectiveness they did not test. Process support should also help students maintain more coherence between the problem, the perceived causes and the solutions specified.

Logical Framework Analysis

Logical Framework Analysis (LFA) was developed for US-AID in the 1970s to improve project management of development projects. In current use the tools of LFA are embedded in more cyclical project management approaches as ZOPP (Deutsche Gesellschaft für Technische Zusammenarbeit (GTZ), 1997) or Project Cycle Management (European Commission, 2001; Team Technologies Inc, 1999) that are used by organizations as EU, Worldbank and OECD to define, manage and evaluate projects and programs (Sartorius, 1996). LFA applies a number of steps, tools and techniques from which we only discuss those we have used.

LFA contains a number of steps from problem analysis to problem definition, with each step leading to a defined product. The problem analysis results in a cause-and-effect representation, called 'problem tree'. An example of such a problem tree for the problem of high infant and material mortality rates is given in Figure *5.1*.

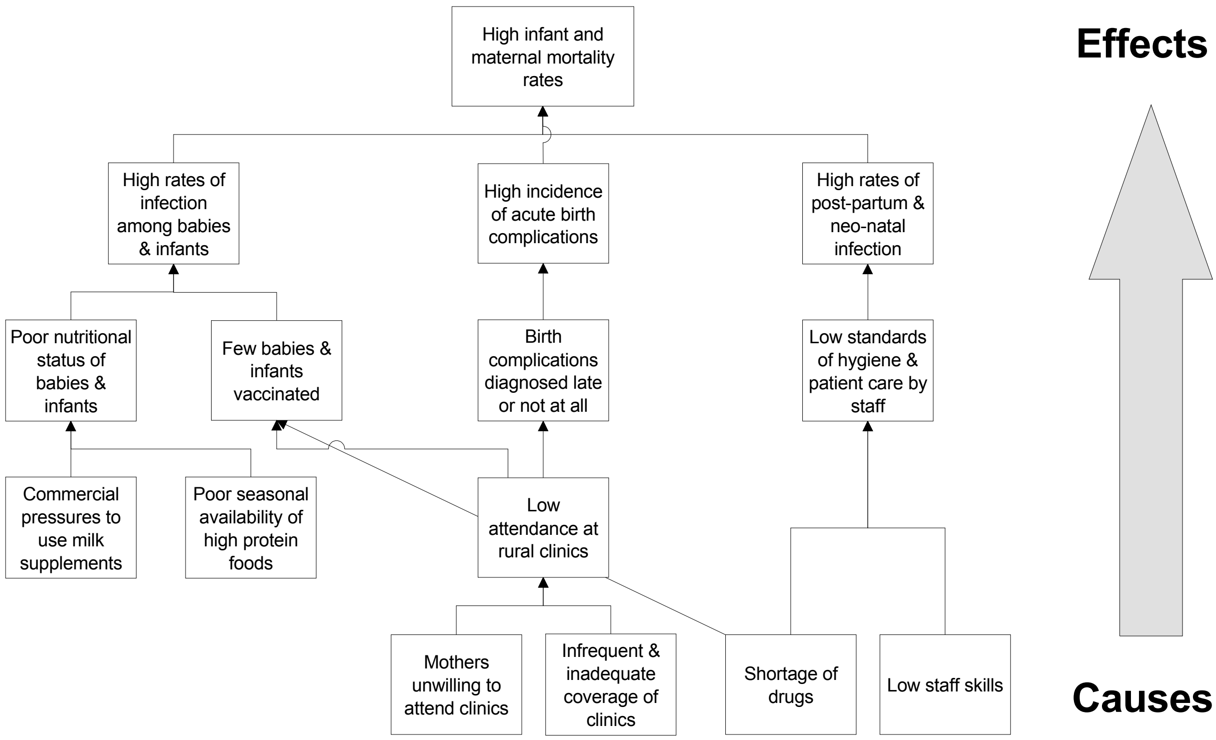

Figure 5.1. Problem tree adapted from EU project cycle management training handbook, Brussels, May 1999.

The components of this tree are then reformulated as objectives, subgoals to be achieved in order to solve the problem. The objectives tree corresponding to the problem tree reported above is presented in Figure 5.2. As can be seen one of the causes of infant mortality 'birth complications diagnosed late or not at all' is reformulated as the objective 'increased/earlier diagnosis of birth complications'.

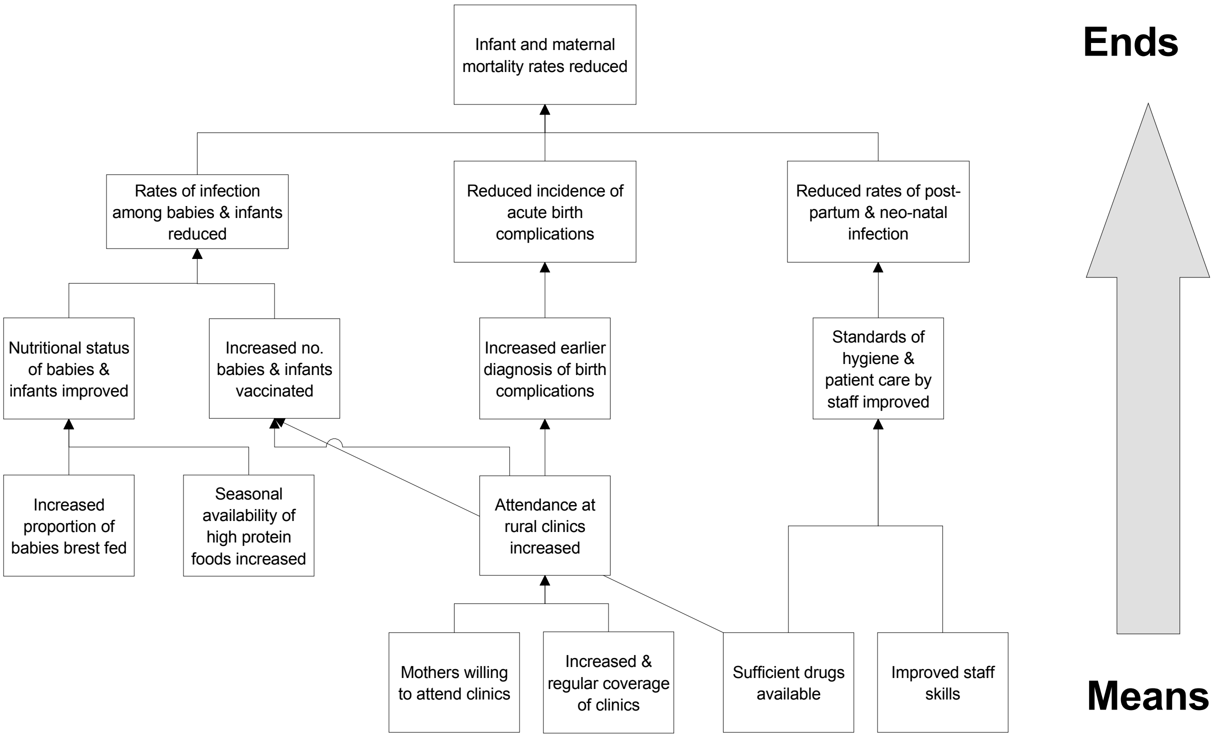

Figure 5.2. Objectives tree adapted from EU project cycle management training handbook Brussels, May 1999.

A next, important, step is to make a distinction between the objectives that the project can attain (the IN-objectives) and those that it cannot attain (the OUT-objectives). The 'IN-objectives' are grouped so as to decide which common interventions can be defined to reach the objectives. Eventually a so called 'logframe' is constructed. This is a matrix that specifies the interventions, the expected short-term, mid-term and long-term effects and how to measure these (performance indicators) and the external (risk) factors, called 'assumptions'.

As can be seen, LFA scaffolds the use of particular types of information. In the first place LFA emphasizes the difference between the (surely attainable) short-term goals and the assumptions that come into play before any long-term goals may be reached. Other features, in particular the use of performance indicators and means of verification, may have less effects on the process but still guide the users towards particular solutions, because they hint at constraints for the solutions.

There are shortcomings to LFA (Eggers, 1998; Gasper, 2000) that have to be considered here. We will limit ourselves to the shortcomings relevant to the current context. First, LFA emphasizes the importance of the logical nature of the analysis and resulting project design, but it offers no means for expressing arguments that support the *causal* decomposition of the problem, the objectives or the interventions. Although LFA guidelines emphasize the importance of the 'project logic', and especially of the causal relations between interventions, short term and long-term effects, they do not force participants to explicate the 'theory' behind the project, as in theory-driven evaluation (Chen, 1990; Rossi et al., 1999). From the perspective of evaluation one may assert that LFA endorses the idea of a project theory without offering the proper means of formulating such a theory. This makes the approach vulnerable to what Gasper (2000) has called 'lack frames' where the logic is lacking. Second, the constraints for acceptable solutions are not represented in one place but, if expressed at all, are dispersed over stakeholder analysis, performance indicators and assumptions, and are (implicitly) applied at the moment the IN-objectives and project strategy are determined.

The study to be reported here is a second step in testing the representational notation. It is a small-scale and exploratory study.

Design and methods

Participants

The participants of this study were sixteen students enrolled in an International Masters Program on European Public Affairs (EPA) at the University of Maastricht. The students came from a wide diversity of national and educational backgrounds. Most students had completed Bachelor Programs in Public Administration, International Relations, Politics, History and Art. There were no Education students in the teams, but there was one Psychologist. The students were confronted with the same problem of reduction of school drop-out with a few changes to make it applicable to the European Union context. Before the experimental session, the students were trained by the author in the basics of LFA. The data were collected in a session that followed a full day of training in core elements of LFA and a two week period in which the participants had worked in four teams and used LFA to define a project on traffic safety (to which they had not yet received feedback). The session was followed by half a day of training with feedback on their assignments and discussion on some of the weak points of LFA presented above.

Design

The four existing teams were instructed to compile recommendations for the Commission of the European Union on efficient policies to reduce school drop-out in the countries of the European Union. They were instructed that they would have to present their results within a month to representatives of the Commission and that they would have only one more meeting before this presentation. Finally, they were told that they could formulate research questions to be answered before their next meeting. The instruction was formulated so as to match the one from the experiment reported in the previous chapter.

Two teams (B and D) were instructed to use Logical Framework Analysis and to create a problem tree and an objectives tree on the wall of the experimental room using colored Post-its. The teams were also instructed to decide on the IN and OUT-objectives. They were discouraged to produce a detailed logframe, because we expected that working out the details of, for instance, the performance indicators would be too time-consuming.

The two other teams (A and C) were told that they were free to choose the method they preferred and they were instructed as well to use the Post-its to represent their analysis and solution.

Scoring

The four 1-hour sessions were videotaped and the contents of the discussions were transcribed by a typist and independently coded by two trained coders. The scheme used to code the protocols was slightly adapted from the one used in the original study. The most important additions were the categories ‘Planning’ to code for dialogues in which students discussed the tasks and actions to be taken and 'Intervention strategy' to code the segments that were dealing with types or classes of interventions rather than with individual interventions. As in the first study all statements relating to a particular topic, including critical remarks or questions were coded the same.

The typed protocols were initially segmented on the basis of the *turns* in the dialogue. If the coders agreed that a turn needed further segmentation, the turn was split in segments (this affected less than 5% of the turns). Both coders were unaware of the backgrounds of the participants or the particular instructions of the groups[8]. The coders were instructed that, only in cases where the typed protocols contained explicit notes on intonation, they could use these notes to derive the meaning of interspersed affirmatives, negations etcetera. We have not tried to have the coders agree on the coding of the segments to get uniform data for subsequent analyses, because that would only suggest better reliability. We decided to base the subsequent analyses on average values over the coders.

The representations that the students prepared using the colored Post-its were transcribed for analysis using flowcharting software (see the example below).

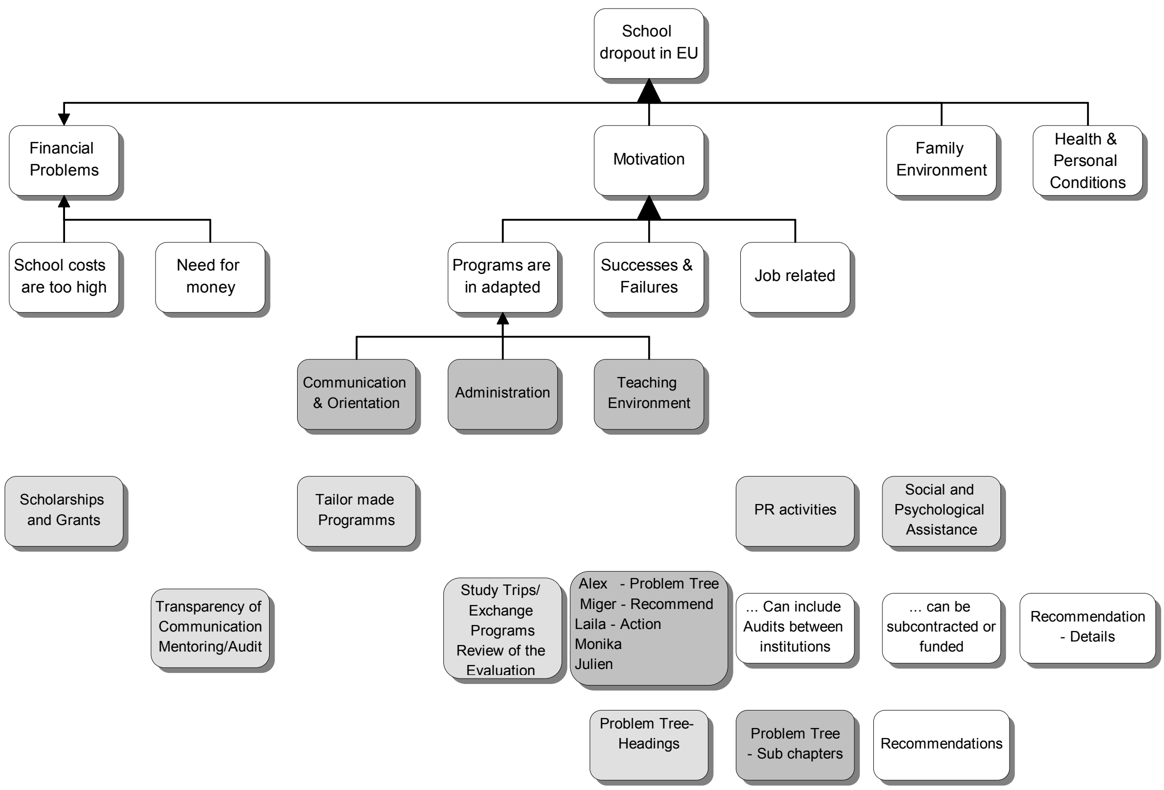

Figure 5.3. Group C's analysis and solution.

[8] Coding and analysis of the protocols were done with the MEPA program made available to us by dr. Gijsbert Erkens of the University of Utrecht.

Expectations

If these students follow the LFA steps in analysis, definition of objectives, and selection of interventions they will demonstrate a more evenly distributed pattern of the sort of topics they address than we found by the previous groups, who tended to concentrate on particular topics. We also expect these protocols to show a clearer division in stages: analyze the problems, state the objectives, select objectives, as described above.
We have also noted weak point of LFA and we expect to see that mirrored in topics that will receive less attention in the current protocols. For example, LFA emphasizes building a problem tree but does not pay attention to its underpinning – that is data presented as evidence for causes. There is no obvious way in which constraints to a solution can be expressed in the LFA structures, that is they could be expressed in stakeholder analysis, but that was not part of the training. LFA emphasizes the outcomes of the interventions and we expect those to be visible in the dialogues.

Results

Approaches adopted

As stated before, groups A and C were instructed that they were free to select any method they preferred and did not have to use LFA. Group A adopted a target group method that one of their participants had worked with before. In this approach groups at risk (target groups) are identified as well as programs (intervention strategies) to address these groups and the persons and/or organizations (actors) that have to be involved in executing the programs. Group C adopted some LFA notions and produced one representation (see *Figure 5.3* above) that contains a 'problem tree'. Groups B and D followed the LFA approach more to the letter and produced a problem tree and an objectives tree. Group D did an attempt to formulate a logframe as well. Both groups did not make an explicit distinction between their IN and OUT-objectives.

Analysis of the protocols

Using the coding scheme the coders could code more than 70% of the segments in the protocols. The agreement percentages on the main categories for the four protocols were 75.3%, 78.1%, 73,4% and 77,3%. Inter-rater reliabilities calculated as Kappa's for the main categories were .65, .69, .63 and .68, which is sufficient for a further analysis of the data. Although the reliabilities found are more homogeneous than in the previous study the Kappa's for the subcategories show a similar variety as in the study reported in the previous chapter. In particular the Kappa's for the categories 'Constraints and Features', and 'Recommendations' are low for three of the four protocols and statements on segments with these codes are to be treated with caution.

The total number of segments and the proportions of the main categories in the protocols of the four groups are reported in Table 5.1.

Table 5.1
Proportion of Main Categories in the Protocols

	Non-LFA instruction		LFA instruction	
	A	C	B	D
N(segments)[a]	454	545	452	547
Analysis[b]	29.5	33.6	38.5	26.0
Constraints	2.6	0.5	0.3	1.3
Intervention	36.9	28.9	25.8	41.0
Planning	4.9	5.9	6.2	5.6
Recommendation	0.2	1.6	1.7	0.1
Research Question	0.0	0.1	1.4	0.8
Other	25.9	29.5	26.1	25.2

[a] Segments are absolute numbers
[b] 'Analysis', 'Constraints' … 'Other' are percentages

Table 5.1 shows that the number of segments of the groups is less diverse than was the case with the beginning experts in the previous study (where they ranged from 250 to 921). The frequencies of segments coded as 'Other' ranged from 25% to 30%., whereas it ranged from 45% to 50% in the protocols of the beginning experts of the previous study. More of the segments could be coded as being 'on-task'. Although two new coding categories were introduced, they alone cannot account for the difference. The new category 'Planning' was used for 5 to 6 % of the coded segments. The new category 'Intervention Strategy' was coded as 'Intervention' in the previous version of the coding scheme. Our conclusion is that the four EPA groups worked more on-task as compared to the beginning experts.

As was the procedure in the previous study, we have rank ordered the proportions of the main categories, excluding the segments coded as 'Other' to get a clearer picture of the relative importance of the main categories. The rank orders presented in Table 5.2 give a very homogenous picture of all four groups. The order of the most important categories is very clear: 'Analysis' and 'Intervention' on ranks 1 or 2, followed by 'Planning'. Then there are slight variations. Group A then addresses 'Constraints' (Features and Risks) – it reaches rank 4. As we indicated above, constraint and features are aspects of a project that cannot be expressed easily using LFA notation. In the rank orders of the LFA groups this category reaches rank 4, 5 and 6. Recommendations are rank 5 for the Target-A group, but reaches 4 for two and 6 for 1 LFA group. Research question is on rank 5 or 6 for all groups.

Table 5.2
Rank Order Scores of the Main Categories

	Control		LFA	
Categories	A	C	B	D
Analysis	2	1	1	2
Constraints	4	5	6	4
Intervention	1	2	2	1
Planning	3	3	3	3
Recommendation	5	4	4	6
Research question	6	6	5	5

We calculated rank order scores for the minor categories which are presented in Table 5.3. The category 'Intervention' is less prominent, because the coding scheme now contained a category 'Intervention Strategy'. The table shows that there are a few minor differences between the groups. For instance the rank order of 'Intervention Effects' for Group B is lower than in the other groups; i.e. group B spent - relatively - more fragments on the effects of interventions than the other groups. Group A spends most of its fragments on 'Intervention strategy' (1) and concrete interventions (2) whereas the other groups emphasize hypothesis analysis.

Table 5.3

Rank Order Scores of the Groups on the Minor Categories

	Control		LFA	
Categories	A	C	B	D
Analysis Data Authority	10	10.5	10.5	11
Analysis Data Experiential	4	4	3	4
Analysis Hypothesis	3	1	1	2
Analysis Problem Scope	5	5	5	5
Constraints	8	9	10.5	9
Features	7	8	8	8
Intervention Assumption	9	7	9	7
Intervention Effects	11	10.5	7	10
Intervention	2	2	2	1
Intervention Outcome	6	6	6	6
Intervention Strategy	1	3	4	3

The top 5 categories of the groups are represented in Table 5.4. There is a remarkable homogeneity between the groups. In the previous chapter differences between the groups became better visible by comparing the 'top 5' categories of the groups and the categories that were unique to the groups by their presence or absence. The picture for the EPA groups however is very clear: there are no categories that are unique (whether present or absent) to a particular group.

Table 5.4
Top 5 Rank Order Categories

Control		**LFA**	
A	**C**	**B**	**D**
Intervention Strategy	Analysis Hypothesis	Analysis Hypothesis	Intervention
Intervention	Intervention	Intervention	Analysis Hypothesis
Analysis Hypothesis	Intervention Strategy	Analysis Data Experiential	Intervention Strategy
Analysis Data Experiential	Analysis Data Experiential	Intervention Strategy	Analysis Data Experiential
Analysis Problem Scope	Analysis Problem Scope	Analysis Problem Scope	Analysis Problem Scope

In order to compare these results with those presented in the previous chapter we corrected the data to handle the differences between the versions of the coding scheme used. For the current groups, we refer to the as EPA groups, we joined the categories Intervention Strategies and Intervention and dropped codes related to the presentation of the recommendations. For the beginning experts the code category Principle was dropped and we removed codes related to the preparation of the presentation. For reasons that will become clear when we discuss the data, we present the data in Table 5.5 sorted on the rank order for the first LFA group (LFA-B).

Table 5.5
Rank Orders for the Minor Categories for the Beginning Experts and the EPA Groups

	Beginning Experts[a]			EPA groups			
	P	E	D	A	C	LFA-B	LFA-D
Intervention (-strategy)	1	1	1	1	1	1	1
Analysis Hypothesis	8	3	3	2	2	2	2
Analysis Data Experiential	6	2	2	3	3	3	3
Analysis Problem Scope	4	5.5	4	4	4	4	4.5
Intervention Outcome	6	7	6	5	5	5	4.5
Intervention Effects	9	9.5	7	10	9.5	6	9
CF Features	2	4	8	6	7	7.5	7
Intervention Assumption	6	8	5	8	6	7.5	6
CF Risks	3	5.5	9	7	8	9.5	8
Analysis Data Authority	10	9.5	10	9	9.5	9.5	10

[a] Beginning expert groups are Philosophers (P), Educators (E) and Debaters (D)

Inspection of the rank orders for the EPA groups shows a number of homogeneous results. Except for one tie, the rank orders 1 thru 5 are the same. For the EPA groups C and D this even holds (considering ties) for all rank orders. The remaining differences between the EPA groups are very small and they deal with a small percentage of the original coding. Ranks 1 to 5, which are the same for the EPA groups, cover 95.5% to 98.4% of the on-task coded segments in these protocols.

If we compare and contrast the beginning experts with the EPA groups we could repeat the observations made in the previous chapter on the unique characteristics of the Philosophers' approach. More interesting, however, is the comparison between the Educators and Debaters and the EPA groups. The first three categories ('Intervention', 'Hypotheses', 'Data Experiential') are shared, with the EPA groups having hypotheses at rank 2 and 'Data Experiential' on 3. Although this is a small difference in rank it is considerable in terms of the proportion of the dialogues. The proportion of the dialogues that the EPA groups spend on hypotheses ranges from 21% (A) thru 29% (B) and 30% (D) up to 37% (C). In the previous chapter we reported for the Educators and Debaters proportions of 10% and 14%. These values however are underestimated, because the Educators, in particular, spend 17% of their dialogue on the preparation of their presentation. Nevertheless, even after a correction for this underestimation the difference remains substantial. It is also what we would expect, because the problem tree of LFA focuses students on the causes of the problem. With the exception of group A, we would expect that hypotheses are often mentioned in the dialogues.

The results for the dialogues on Data are also as we predicted. In the previous chapter we reported that the Educators and Debaters spent (underestimated values of) 24% and 20% of their dialogues on 'Authority' and 'Experiential Data'. For the EPA groups we find 14% (Target-A), 17% (B), 9% (C) and 5% (D) of the dialogues spent on 'Data'. This is what we expected for the LFA groups, because LFA does not encourage specification of the underpinnings of the problem analysis. The result also demonstrates that improvements can be made here.

Qualitative analyses

The position of the EPA students is somewhat different as compared to that of the beginning experts of the previous chapter. First, the EPA students are from different national backgrounds and not all of them are fluent in English. In several episodes in the protocols the students are trying to find the correct wording, or they explain formulations to each other. Second, they come from countries with different educational systems. The protocols contain several episodes in which the participants share information on the educational systems and regulations of their countries. Finally, the recommendations that they compile cannot assume a uniform educational system, because the EU member states have different educational systems.

The protocols all contain segments where the participants try to delineate what drop-out is. The participants try to define drop-out in terms of the level of schooling at which a learner leaves the educational system, that is where does leaving school count as drop-out. This is a rather shallow understanding of drop-out and why it is a problem. There is only one statement in the protocols that indicates a deeper understanding of the problematic side of drop-out: “It’s that you don’t have the right education to actually start any kind of career” (protocol A, segment 61).

Groups B, C and D all tried to create a problem tree to represent the drop-out problem. Inspection of the trees learns that, rather than cause-and-effect chains, as LFA would have it, these trees have abstract problem categories, such as ‘family problems’ or ‘social environment problems’ directly connected to the major problem, that is drop-out. Below the abstract categories we find the more concrete causes (‘drug abuse’, ‘gangs’, ‘teenage pregnancy’). Note that this is *not* the structure one should define when using LFA. The problems tree in the example reads (according to LFA) as “drug abuse, together with gangs and teenage pregnancy, *causes* social problems; social problems *cause* drop-out”. The use and position of the abstract problem categories will lead to problems in the next step of LFA, the formulation of objectives. According to LFA, there is a one-to-one correspondence between the elements of the problem tree and the elements of the objectives tree. Thus, the abstract problem categories have to be rephrased as objectives. We would then find objectives, such as ‘less social problems’ that have to be met before the objective ‘less drop-out’ can be reached. The LFA style is to either only use more tangible problem and objective statements, or put the general categories *below* the concrete ones. As we will see below, the groups that prepared an objectives tree found different solutions to this LFA-based problem.

In the following discussion we concentrate on the protocols of the LFA groups, but we will start with a brief overview of the protocols of groups A and C. *Group A* is the only group that did not create a problem tree. This group decided to identify the target groups that were at risk of dropping out and then identify the interventions and the actors responsible for them. This is a method one of the participants was familiar with. Their protocol reads as a lengthy systematic brainstorm in which various target groups are identified. Working down the list of target groups they described interventions as well as the actors. *Group C* developed a problem tree, but then decided to not develop an objectives tree: “I don’t think that we need to do objective tree in the moment (…) because we discussed we can just develop recommendations" (segment 304). They return to the issue at the end of the protocol when someone suggest to build the objectives tree: "You can try to make the objectives tree. I mean it is quite easy to" (segment 541), but someone concludes: "We can do the problem tree (sic! JvB) but for us it was much easier to see what is the problem and then think about the recommendations" (segment 544). Group C also decided that they would address only two of the four major problems identified ('financial problems' and 'motivation').

Group B was instructed to apply LFA. They start to delineate the problem by discussing at which educational levels drop-out may occur. They then try to address the causes (segments 25, 26) but get sidetracked by a new discussion on the meaning of drop-out and the educational levels (segments 29-41). In segment 42 someone proposes to address causes followed by the suggestion that 'insufficient entrance requirements' to high school is a potential cause. This however leads to exchange of information on entrance requirements to university (segments 46-60). New exhortations to concentrate on causes (segment 61) and to brainstorm (65) lead to a few minutes during which most of the components in the problem tree are formulated. Their analysis is in terms of abstract categories, such as 'social problems' and 'family problems'. Concrete, isolated causes are suggested in this phase, but not accepted "(…) we should first stick to the bigger causes like family problems".

From segment 173 on, eighteen minutes in the session, group B starts to work on the objectives. They have used two general problem categories 'drawbacks of the educational system' as well as 'social problems' in the top of their problem tree. They now have to reformulate these as objectives. They notice the problem that the abstract categories are causing, and react with humor: "So improved educational system … wild". "And their salvation of all social problems in the world" (segments 192-193). They formulate an objective for 'social problems'. "It sounds so ridiculous: reduction of social problems" (segment 198). Eventually, they agree on the formulation 'assistance to victims of social problems' (segment 203). Their objectives tree thus becomes a mixture of objectives and interventions (see Figure *5.4*). Eventually, they decide on a subset of interventions that can be addressed in short term as their IN-objectives.

Lengthy passages of the protocol are filled with two of the four participants writing out the problem tree and the objectives tree, which were used to capture the results of their work.

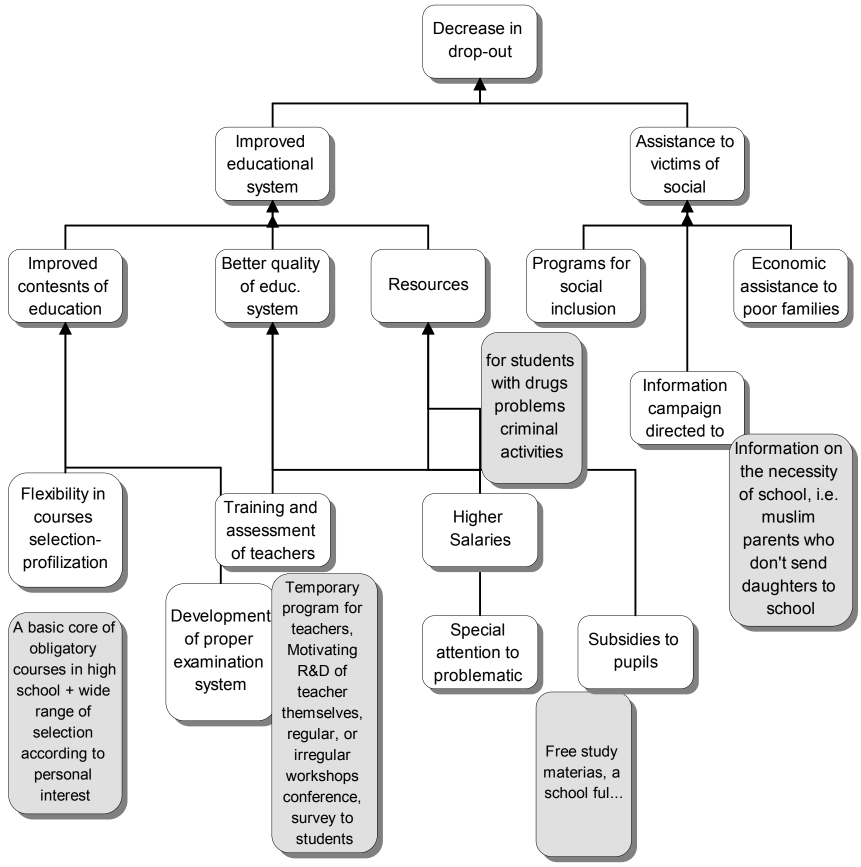

Figure 5.4. The objectives tree of group B containing objectives as well as interventions. The recommendations are in the grey boxes.

Group D was also instructed to apply LFA. Their protocol starts with a very brief orientation on the task and the meaning of drop-out, before the group starts to brainstorm in segment 7. Within the first five minutes of the protocol most of the content of the problem tree is already mentioned. Again, we see abstract problem categories appear as top nodes: 'family problems', 'social environment problems' and 'problems at school'. The brainstorm of the group is interspersed with suggestions of possible interventions, such as policies against child labor, establishing a psychological counselor at schools, improve second chance and evening schools and others. When this group has to address the objectives (late in the protocol – at 55 minutes) they decide to drop a number of problems from further consideration: "(…) for the family problems we cannot give objectives. They are not amenable to change" (segment 421). From then on, they loose the relation between problem and objectives completely by adding several interventions to the objectives tree that are not related to either problems or objectives.

From the protocols of both LFA groups as well as from their products, the problem trees and objectives trees, it is clear that the groups failed to apply the LFA techniques correctly. In particular they failed to maintain the direct links between the formulation of problems and objectives. There are some other, informal, indications that this is not an incidental failure. Before the problem solving session each group had prepared a small project proposal in the area of traffic safety that was to be presented during the final session. Group D, for example, had addressed the safety topic from the perspective of 'driver psychology'. In their problem tree three categories of major problem areas are present right below the top of the problem tree: 'perception, cognition and motor skills'. Similar uses of abstract categories below the top of the problem trees can be found in the assignments of groups B and C. Only group A prepared a problem tree that consisted of cause-and-effect chains only. The data from these assignments suggest that the patterns that we observed in the protocols are based on more persistent misunderstandings of certain LFA techniques, which leaves room for improvement of the LFA course that these students followed.

More interesting than the mastery of LFA techniques is how our participants approach the problem presented to them. In chapter 2 we discussed research by Voss, Tyler, et al. (1983). They described how experts when solving social science problems spend a considerable amount of time to build a problem representation, delineate the problem constraints and develop an orientation as to the type of problem, for example a political or a technological problem. Their solution (as well as diagnosis) is abstract and only a few solutions are considered. Novices on the other hand mention separate, concrete causes to which they directly connect solutions.

Our participants are not experts. None of them had particular expertise in the area of drop-out, so all teams had to rely on there own insights and experience. None of the teams formulated a solution that could be compared to the 'charter school' that the Educationalists came up with in the previous study. This, however, does not imply that they operated as novices. Groups B, C and D do not approach drop-out as a simple problem having one single cause or a number of unrelated causes. Instead, they use abstract categories of causes from which they try to derive more concrete causes. They devote considerable attention to the causes (more then the beginning experts did). Although they run into trouble applying LFA, at least one of the LFA groups (group B) managed to maintain at least some of the relations between the problems identified, the objectives and the interventions. Group A followed a different approach by identifying various groups at risk, as such their analysis may look closest to that of the Debaters who identified a series of unrelated causes. The important difference is that group A did not consider the target groups in isolation. When they were defining their interventions they were combining interventions as well as target groups.

All this does not correspond to the novice behavior described by Voss et al. (1983). Yet, the approach is also quite different from the expert problem solving approach. Whereas the experts take care to delineate the constraints to the problem, we find, as already noticed in the quantitative analysis, that our participants hardly address constraints (less than the beginning experts, in fact) except for the question at which educational level drop-out may be located. The instruction of the participants did not offer any clear constraints but only told them that the Commission was seeking their advice in developing 'effective policies', which is a statement that leaves room for different interpretations. The protocols contain no statements in which the participants are discussing this constraint. Neither can we find statements where they are applying it as a criterion to evaluate interventions or recommendations. Group C and D define their own constraints when they decide to not address some of the problems that they have identified, because they consider them as being not amenable to change. Finally, the protocols contain very few statements about the effectiveness of recommendations or interventions in reducing drop-out, a point that we already mentioned in the quantitative analysis.

All in all, the EPA groups present a mixed picture of expert-like and novice-like behavior. In particular the lack of attention for constraints and effectiveness of recommendations is a reason for concern.

Discussion

In this chapter we analyzed the problem solving protocols of four multi-disciplinary groups of students that were trained in the basics of Logical Framework Analysis (LFA). The groups worked on a similar school drop-out problem as was used in the previous chapter. We found no such pronounced approaches as were identified for the Philosophers and Debaters in the previous chapter. Two reasons might explain that. *First*, our teams were multi-disciplinary and this alone may have been sufficient to prevent the sort of mono-disciplinary approach followed by the Philosophers and the Debaters. *Second*, although none of the groups had specific expertise in the area of school drop-out, all participants brought the common experience of the EPA program that deals with European Policies and most participants had educational backgrounds that confronted them with social problems, policies and interventions. Therefore our participants, compared to the Philosophers and the Debaters, may have brought more relevant knowledge to the session than just the LFA training.

We found that the LFA groups spent considerable effort discussing the causes of the problem – as one would expect given that they all started to develop a problem tree. They expressed these causes using abstract problem categories, whose direct relation to the problem of drop-out is not always clear, however. They used abstract descriptions of the interventions as well. The new category 'intervention strategy' received a high ranking (see Table 5.3).

We found that all groups paid little attention to delineating the problem constraints and to formulating the underpinning of their analysis and solutions. As we mentioned earlier this corresponds to a lack of expressibility in the data structures of LFA. Finally, we noted that the LFA data structures were used as a means to consolidate the results of deliberations, rather than as representation of an ongoing dialogue.

The most important question to be answered here is whether Logical Framework Analysis offered the process support that we considered necessary for the non-expert groups in the previous chapter. Consider how the EPA groups worked in comparison to the beginning experts. Three of the four groups tried to apply some of the LFA techniques, and group A also worked along systematic lines by applying a target groups approach. None of the groups produced solutions comparable to those of the Philosophers or the Debaters. The EPA groups more or less ignored constraints, to which the Philosophers devoted their attention almost exclusively. The EPA groups did not formulate a number of unrelated causes and how to implement measures to eliminate each of them, as did the Debaters. The beginning experts, in particular the Debaters, ran into some problems when they came to realize how the social nature of some of the causes would prevent direct solutions. The EPA groups quickly dismissed a number of problems as being not amenable to change. Of course, their analysis of school drop-out is shallow compared to that of the Educators, which makes it easier to dismiss complete problem areas.

We found no evidence that the two LFA groups produced deeper analyses or more coherent recommendations than the other two EPA groups. In fact, the quantitative analysis already indicated that the differences *between* the LFA groups were considerable. Nevertheless, we may say that LFA at least seemed to guide our participants in addressing the problem in a number of separate steps. The groups try to first analyze the causes of the problems before they work out interventions.

More important however is the guidance effect that the representational notation of LFA seems to have on the contents of the dialogues. The ontology of LFA addresses causes and their relations (problem tree), objectives and their relations (objectives tree) and interventions, outcomes, effects and risks (logframe). We see that reflected in the dialogues. As usual, most of the segments are spent on interventions and hypotheses, i.e. causes. In fact, the percentage of the LFA groups is higher than what we found for the beginning experts in the previous study. Fewer fragments, also less than we found with the beginning experts in the previous study, were spent on data. Constraints, assumptions and risks are hardly mentioned. So, although we find some indications of process support offered by LFA, it as well does not succeed in guiding students to address the structural relations between the components of the analysis and the solution.

Conclusions

All in all then, LFA could be used to offer a limited type of process support. There are a number of serious drawbacks however. *First*, the methods and techniques of LFA are not self-evident, implying that we have to train students in the use of them. And this would obviously require more training than the EPA students received in order to prevent the common errors that we encountered in the formulation of the problem and objectives tree. There are, however, other serious drawbacks associated with LFA that need to be considered.

Second, the ontology of LFA is limited. Although LFA emphasizes the importance of *logic* in a project definition, the method seems to lack the means of ensuring that the data-structures produced indeed have this quality. LFA offers no means to represent (and therefore does not encourage to express) the underpinnings of the relations in its trees (that is evidence that A causes B, and that Intervention I will lead to a particular result). One might expect this to occur with novices but Gasper (1999; 2000) made similar observations about the use of LFA in international programs. In a similar way LFA does not allow or solicit the expression of the constraints to the problem, and we see that in the EPA protocols constraints as well as underpinnings are rarely used.

Third, the usability of LFA may be related to the level of domain expertise and the context in which it is used. If LFA is used in project development a group of experts may be involved in defining the problems and objectives tree. Their joint ordeal *may* be sufficient for further project development.

In educational settings, however, that level of expertise cannot be assumed. Moreover, we often are concerned more with students demonstrating careful analysis with sufficient backing and argumentation underlying a solution than with a 'correct' solution. Considering these drawbacks and the limited guidance that LFA seems to offer we conclude that LFA does not need to be incorporated as guidance and support component in the representational notation.

6 Towards computer-supported argumentation visualization for social science problems

The work reported here started with the claim that social science problems present to students the type of open, authentic tasks needed in the competency-based education that the Open University of the Netherlands is aiming to develop. The question that we sought to answer in the studies reported here was how collaborative solving of this type of problems, can be supported with the use of external representations. Given that the analysis, solutions and evaluation of solutions of these problems are processes involving extensive argumentation, we concentrated on graphical representations of these arguments. One of the most important questions to be answered is what representational notation is needed to support collaborative solving of social science problems. That is, what are the objects and relations that we offer students to represent the argumentation with?

The ultimate goal of the research is to offer students support and guidance in their collaborative problem solving activities through Computer-Supported Argumentation Visualization (CSAV). The development of such a CSAV environment will take a considerable amount of time (and money) and here we have attempted only to answer some of the initial questions in the design of a CSAV environment. The steps that we took can be mapped onto the six stages of the development of CSCL environments described by Kirschner (2002). His stages are inspired by the observation that CSCL environments need to be carefully designed and developed if we want to ensure that the support contained in these environments is perceived and used (as intended) by the learner. Kirschner discusses the following stages in the development of CSCL environments:

1. Determine what learners actually do or what they want to do
2. Determine what can be done to support those learners
3. Determine the constraints of the learner, learning situation and learning environment and the conventions that already exist
4. Determine how learners perceive and experience the support that we provide
5. Determine how the learner actually uses the support provided
6. Determine what has been learnt

The work reported here can be described as concentrating on the first three stages of the development cycle. Chapter 2 develops the model to describe what learners do and how external representations may support them. Chapter 3 can be considered an intermezzo in which we learned that adding-on to an existing CSCL environment does not work. Chapters 4 and 5 are studies on how learners go about solving social science problems. They are intended as a very first step towards testing the usability (and usefulness) of the representational notation developed in chapter 2. In chapter 5 we tested Logical Framework Analysis as a means for guidance and support. Here, the conclusions of these studies are translated in a first set of global functional requirements for a CSAV environment. Before presenting these global requirements the next section of this chapter reviews the results of the studies reported in the previous chapters and draws a number of conclusions with respect to functional requirements and further work.

Review of the results

In the previous chapters we reported a number of explorative studies aimed at improving our understanding of how we can (or cannot) use, in a computer-based environment, external representations of argumentation to support collaborative solving of a type of ill-structured problems referred to as social science problems. In these explorations we have used results from different areas, including problem solving research, evaluation research, research in Computer Supported Collaborative Learning (CSCL) and project methodologies, in particular Logical Framework Analysis.

In chapter 2 we reviewed the literature on solving social science problems and 'wicked problems'. These concepts were developed in different research areas and we attempted to synthesize them in a multiple agent, multiple representation model. The different traditions have also inspired different lines of CSAV environments. We reviewed a number of CSCL environments, that emphasize CSAV for scientific inquiry (addressing 'why questions') and a number of design-based environments that are more focused on the 'how questions' and enumerated a number of different services found in these environments. They all offer representational services, that is, they have functionalities with which the users (or a mediator) can manipulate a graphical representation. CSCL environments offer communication services by which learners can communicate in synchronous or asynchronous mode. Finally, we identified guidance and support services and we noticed that these services often are dependent on the representational service. Whereas discussion-based tools, such as CSILE®, often hide information in the text of discussion nodes, guidance and support services need to have access to information such as plausibility ratings and ratings of agreement.

In chapter 2 we presented a model for problem solving using three states – orientation, solving and evaluation – and a number of cognitive and communicative demands related to these states. These demands were then related to the services offered by CSAV environments. Finally, chapter 2 described a first version of a representational notation to support collaborative solving of social science problems. The claims made in chapter 2 are summarized below. As can be seen several of them deal with more general descriptions of the type of problems and problem solving processes, such as statements 1 thru 5. Others, are more directly related to the services we identified. We will reconsider these claims as we review the studies reported in this thesis.

1. There is no unambiguous statement of an ill-structured problems or its solutions (Reitman, 1965)
2. Social science problems are ill-structured problems with 'delayed evaluation' (Voss, Tyler, et al., 1983)
3. Solving of social science problems involves informal reasoning to establish the goals, constraints and solutions to the problem (Voss, Greene, et al., 1983; Voss, Tyler, et al., 1983).
4. Solving of social science problems involves different stakeholders (Kunz & Rittel, 1970)
5. Stakeholders bring their own viewpoint to, i.e. representations of, the problem (Kunz & Rittel, 1970; Rittel & Webber, 1984)

6. Different agents / problem solvers may have shared, partially shared and private representations of the problem and procedures (Alpay et al., 1998)
7. The representations of the agents may be temporary or permanent: the latter coincide with systems, procedures and models used on a regular basis (Alpay et al., 1998).
8. Problem solving consists of three states: orientation, solving and evaluation (Newell & Simon, 1972)
9. Each problem-solving state has its own representational and communicative demands (Duffy et al., 1998; Sloffer, et al., 1999).
10. Evaluation of solutions is (also) based on the accuracy, consistency and plausibility of the underlying argumentation (Voss, Tyler, et al., 1983)
11. Maintaining accuracy, consistency and plausibility are important functions of representation management (Alpay et al., 1998)
12. The criteria by which accuracy, consistency and plausibility are evaluated need not be the same for all agents / problem solvers (Voss, Wiley, et al., 1999)

Chapter 3 reported a study, conducted at the University of Ghent, in which we used an existing CSCL environment, Belvédère®, that was designed to support collaborative scientific inquiry. We varied the level of detail of the meaning of symbols to support deliberations on concrete actions to solve a problem. This experiment suffered from a number of technical problems leading to loss of data that could have given us a better insight in the problem solving activities of the student groups. On the other hand, other technical problems forced us to introduce a 'condition' to the experiment in which students worked side-by-side and this condition revealed a number of interesting results. Overall, the results of this experiment were clear-cut: whereas we predicted a number of positive effects of the more detailed representational notation, the opposite occurred. Students who shared one computer, working side-by-side, and using a low detail representational notation outperformed those who were working at separate computers using a high detail representational notation.

These results seem to point to an interaction between the task, the representational notation and the communication and co-ordination facilities that Belvédère® offers. Veerman (2000) found that students using Belvédère® while working at separate computers had to spend considerable effort, about half of their messages in the Belvédère® chat box, on technical issues, planning and coordinating actions on the diagrams. What this suggests is that the communication services offered by Belvédère® do not match the coordination demands that students have. Belvédère® is lacking coordination mechanisms, such as chats that are anchored in the diagrams, or a topic-based agenda that indicates existing or new items in the diagrams that need to be worked on. The data is sparse however and their interpretation is not clear-cut. Therefore a replication of this experiment with balanced conditions is needed before we can make any specific claims as to the nature of this interaction.

From the Ghent experiment we learned that we could not expect that 'transplanting' external representations to Belvédère®, or any other system, would offer learners an environment with which they could learn to solve social science problems using external representations. This was a clear sign to reconsider the design of an environment intended to support such a complex process as the collaborative solving of ill-structured problems. In particular we had to put more emphasis on the initial analysis stages of the design. From this point on we focused on the development and testing of a representational notation, rather than attempting to design representational tools.

In chapter 4 we reported on a first study on how students solved a social science problem. Protocols of three mono-disciplinary groups were made available to us by prof. dr. Tom Duffy. These protocols were collected to study the problem solving activities of groups of beginning experts in different domains, namely Education, Philosophy and Debating. The Educators were considered beginning experts in the content area (school drop-out). The Philosophers were considered beginning process experts in systematic analysis and argumentation and the Debaters were considered beginning process experts in systematic debate and persuasive argumentation. The groups were instructed to compile recommendations to the Board of Governors as to how school drop-out in the U.S.A. could be reduced.

We analyzed the protocols of these three groups using a coding scheme that was derived from a first version of a representational notation presented in chapter 2. The representational notation is an eclectic blend using results of several research areas. Using the coding scheme we were able to analyze a substantial amount of the segments in the protocols, which showed that the students used several of the concepts included in the representational notation. This is encouraging as far as the usability of such a scheme is concerned.

More importantly, the coding scheme proved robust enough to cater for the different approaches that the three groups followed when they tried to solve the problem. Whereas the ‘real experts’, the Educationalists in this case, came up with a reasonable solution, although ignoring several other possibilities, two other groups that did not have an affinity with the problem did not come to coherent recommendations. One group (Philosophers) tended to focus on the constraints of the problem only, while the other (Debaters) worked almost mechanically along a list of unrelated causes that they had previously identified. With the coding scheme these different approaches as well as the shortcomings of the approaches could be typified, thus indicating where further improvements could be made.

Before we state our conclusions from this study we need to consider a number of weaknesses in the study. *First*, the empirical base on which we draw our conclusions is, admittedly, a small one (which keeps us close to the tradition of Voss…). Video- and audiotapes of six parallel groups were recovered and made available to us in December 2002, but could not be analyzed in due time to be reported here.

Second, the reliabilities of the coding, expressed as Kappa’s, are typically in the range of .60’s. This may be sufficient to explore the data, the reliability needs to be improved if we want to make more substantial claims on the basis of the dialogue analyses. For our analyses, we have not made coders agree on the coding of segments, but rather based the analyses on either the means of the codes, or on the coded protocol of one single coder. Admittedly, we still could have had coders agree, not to use that agreement as the code, but in an attempt to explore why the coding scheme was sometimes hard to use in coding the protocols.

Third, we may consider a potential weakness resulting from the fact that we analyzed the problem solving dialogues completely from the perspective of the type of content addressed and have not looked in any detail at the communicative functions of the dialogue exchanges. This may seem a reasonable choice as far as one is concerned with the development of a representational notation. We have to address content if we need to decide about the expressiveness of a representational notation. To validate a representational notation or to test its potential usability one has to check whether the intended audience is already using some of the concepts contained in such a scheme, or whether those that are trained in the notation use the concepts in their problem solving activities. For that reason the coding scheme has to concentrate on the content, rather than on communicative actions.

One may argue that we have concentrated less on content than we describe here, because we have failed to differentiate the statements relating to a particular category of content. All statements in a category – for example an intervention - are coded as belonging to that category, irrespective of the nature of the speech act: thus, whether an intervention is stated or questioned or evaluated makes no difference to the coding scheme. This may seem as a serious omission, but the effects will be modest in the contexts in which we use this scheme and in the context in which the scheme is implemented as a representational notation. To start with the latter: in an implementation one does not 'state' an intervention, one adds an 'intervention object' to the external representation. Other objects are used to make further statements about related aspects: risks, expected outcomes, data on the effectiveness of the intervention, et cetera. In the contexts that we have used the scheme we were not interested in the use of particular categories. We are, to put it bluntly, more interested in whether participants cover all aspects of the solution process than in the details of the particulars.

The choice for a content-centered approach is less obvious when we are studying students that are working in collaborative situations. It has become almost commonplace to argue that we should not expect CSCL to lead to dramatic changes in learning outcome and that we may better expect to see changes occurring in the interactions between students. Analysis of the communicative functions of speech acts in the student dialogues has become more or less standard in CSCL research. As an example, the Utrecht tradition can be mentioned. An impressive series of studies has been conducted that all look at the relation between constructive speech acts and argumentatives in the dialogues of students and learning outcomes (in a very wide sense). In a recent overview (Kanselaar et al., 2003) the authors state, however, that the relation is less clear than hoped.

In contrast to that approach, we analyzed the dialogues between students in order to check whether they used the concepts that we introduced in the representational notation. If only for that reason, we had no alternative than to opt for a content-centered coding. Admittedly, we have driven the content-centered coding beyond this point in those studies where the coders not only classified statements in content categories ('an intervention is mentioned'), but also described the content ('the intervention is 'pay if they graduate''). This however was the only way of differentiating between the breadth and depth of the ideas generated, allowing us to differentiate between the repetitive rephrasing of the same content and generating of different content. It certainly helped to better understand the differences between the problem solving approaches of students from different background. Admittedly however, this approach requires that coders themselves invent coding categories and we have not even tried to establish inter-rater reliabilities for this type of coding.

When we attempt to improve the problem solving processes of students it is not by stimulating them to engage in constructive and argumentative actions in general. The notion behind the methods and behind the analyses of the dialogues is more precise. We surmise that that solving this sort of problems involves the formulation and evaluation of the 'program theory' behind the interventions. The objects and relations that we added to the traditional LFA objects correspond on the one hand to a scientific analysis of the 'why' part and on the other to a specification of the 'program theory' for the 'how'-part. To paraphrase Pawson and Tilley (1997) we are interested in having students explain to us beforehand *whether* particular interventions are likely to work in a certain context and what convincing reasons they have to believe so.

From the study reported in chapter 4 we drew the following conclusions. First, the representational notation that we construed was promising, because using the coding scheme that was derived from it, a substantial part of the dialogues could be coded. This indicates that the students already use concepts contained in the representational notation and that is a sign of potential usability. Second, the representational notation proved robust for different problem solving approaches. Using the coding scheme the protocols of all three groups could be coded, and the different approaches of the groups could not only be captured, but typified as well using the categories of the coding scheme. The different approaches could be described as focusing on particular categories while ignoring certain others. This then brings us to the final conclusion. Considering the solutions produced and the biases noted in the approaches of the groups, more guidance and support is needed for the non-domain experts. We may notice however that the groups and in particular the Debaters worked very systematically. Considering the issues regarding the communication services and the agenda function that were raised in the discussion of the Ghent Belvédère® experiment, we can say that the Debaters in the orientation state, brainstormed and decided on the issues (causes) that they would work on in the solving state and then, in their solution state, addressed the causes topically. In other words, what we need is not guidance and support to help students maintain focus on particular categories, but support that helps them to treat the different categories in their structural relations, for example by testing the effectiveness of proposed interventions or by checking proposed interventions against stated constraints.

In chapter 5 we reported on a study in which we tried whether the methods and tools of Logical Framework Analysis (LFA) could be used to guide the process of solving social science problems and the content addressed. Choosing LFA was more or less obvious because we had already incorporated elements of LFA in the notation. Moreover, LFA is a method that is used in the analysis of the type of ill-structured problems we are dealing with. LFA offers a number of tools – trees, matrices – that are to be filled in a series of steps. However, there are also a number of known shortcomings to LFA, which we discuss below.

The participants of the study reported in chapter 5 were students enrolled in an International Master program (European Public Affairs) at Maastricht University. The EPA program is attracting students with widely different educational and cultural backgrounds. There are several advantages to having these students participate in the study. First, they are all advanced students, beginning experts, some of them already own a master's degree. Second, they are from different national and educational backgrounds and we can form multidisciplinary teams. Third, and finally, a part of the EPA program is concerned with policy evaluation and this part includes training in the basics of Logical Framework Analysis (given by the author).

Four groups of EPA students, trained in the basics of LFA, were confronted with the task to formulate recommendations on how to reduce school drop-out in the European Union. Three of the four groups tried to apply LFA and the analysis of their protocols yielded very similar results. The EPA groups tended to spend more dialogue turns on causes than did the beginning experts of chapter 4 (Philosophers and Debaters in particular). In their analyses they used more abstract problem categories ('social problems') than for instance the Debaters did. More abstract categories do not imply a deeper view, however: the EPA groups produced rather shallow analyses with little underpinning (data) of their analysis and solutions. They spent relatively few turns on constraints and desirable characteristics of the solutions.

We noticed that none of the groups that tried to apply LFA managed to maintain consistency within and between the different representations that LFA uses. This can be traced back to some of the particularities of LFA and it does not signify a weakness in the problem solving activities of the participants. On the contrary, one might say that they avoided the typical novice approach of stating several unrelated causes (Voss, Tyler, et al., 1983). Finally, we noted that the external representations in LFA were used as a means to *consolidate* the results of deliberations.

The results of the EPA groups may be related to weaknesses of LFA. One of the most outspoken critics of Logical Framework Analysis is Gasper (1999; 2000) whose central theme is that Logical Framework Analysis is particularly vulnerable to being applied in rigid ways from which little can be learned. Gasper pinpoints his criticism to the *logframe* because he sees that as the core of the method. Gaspers points to several weaknesses, but there is only one relevant to the parts of the LFA methods and tools that we use. Gasper uses the poignant term ‘lack frame’ to describe a logframe that is construed after the project is defined and that is lacking the project logic that is so central to LFA. His criticism is somewhat curious, however, because it focuses on those aspects of the logframe that are derived from the problem and objectives trees defined previously. In our view a more accurate formulation is that LFA neither enforces, nor encourages its users to articulate the project’s logic and the evidence that supports the analysis or solution. There are no means in the LFA representations to express data that corroborate the analysis in the problem tree or the objectives tree.

Furthermore, there are no facilities in LFA for expressing ‘program theory’, an explanation why a set of interventions will ultimately lead to the solution of a problem. Unfortunately, the use of common terms can be confusing here. The logframe contains a column called ‘assumptions’. These are (pre)conditions outside the scope of the project that must materialize for the interventions to work. In our representational notation we refer to them as *conditions* and *risks*. In program theory the term *assumption* is used to describe a set of presupposed intervening mechanisms that link the direct outcomes of an intervention to the long-term effects. These are called *assumptions* in the representational notation. LFA has no object to express this type of assumption.

In the protocols of the EPA students we find hardly any argumentation on the causes of the problem, or on the reasons why proposed interventions are likely to have any effects. As we noticed, LFA does not encourage making this reasoning explicit, there is no representational guidance towards these categories in the external representations used in LFA.

However, one must also consider that expressing argumentation requires more than adding an object to an external representation, it requires domain knowledge. Although our students are beginning experts, none of them had any expertise in the field of Education or school drop-out. They may have found it difficult to apply their particular expertise (Law, International Relations, Economics) to this problem. An indication here is that we found almost no sign of different problem representations that could be traced to differences in backgrounds. Two noticeable exceptions are a Psychologist in group D who more or less persuaded the others in her team to go for psychological interventions, and the participant in group A who had worked with target group analysis and persuaded her team members to apply that method. So, looking back at the claims that we reported earlier, in particular at those dealing with the multiple agent, multiple representation model we must conclude that neither different representations nor operators corresponding to different educational background were visible in the protocols of the multidisciplinary groups. Task characteristics may have contributed to this effect, however. Students did not prepare for the problem solving session and the synchronous mode of operation may have further contributed to they not being able to develop a viewpoint on the drop-out problem where their domain expertise could be used. Future research should investigate under which conditions, including ones in which they can prepare for these sessions, students can be stimulated to articulate their specific views on a problem.

The students of the EPA groups seem comparable to the Debaters of chapter 4, but they do have certain process knowledge, which they try to apply. Maybe that prevents them from coming up with a list of unrelated causes and maybe that helps them to think of feasible interventions, but they lack the domain expertise to produce anything similar to the charter schools policy that the Educationalists came up with, or to check their proposed interventions against constraints such as EU policies. From these considerations we learned that there are important limitations to the usability of LFA as process support, even if we consider advanced students as the main target group.

The study with the EPA students suffers from a number of weaknesses that partly overlap with those of other studies. First, the reliability of the coding remains a concern. Although slightly higher than in the study with the beginning experts, the Kappa's are still not very impressive. Yet, more segments could be coded with a content coding and it seems likely that we can make further improvements by including an abstract category for causes as well.

This study, as did the previous ones, engaged students, most of whom are not experts in the problem domain, in a short problem solving session, typically lasting one to two hours. We confronted the students with an ill-structured problem where they themselves had to define the problem, delineate its constraints, agree on criteria and stopping rules. Moreover, they were required to specify a set of interventions that would eliminate the causes of the problems, along with an argument why the was likely to work. This is a lot to be accomplished within an hour. Only the Education group, reported in chapter 4, came up with an analysis and a solution that contained these elements, but both were constrained to the Educational realm.

Our explorations show us what sort of biases we can expect if we have students who are less advanced in the relevant domain. With the exception of the Education students all groups demonstrated that they focused on *particular* types of content sometimes completely ignoring others. According to the idea of representational guidance (Suthers, 2001) a representational notation is also offering a task structure for the students. In ordinary language: if you provide them with objects and relations they will tend to use them. So it remains to be seen if our students when they are using a full-fledged representational notation will manage to use the structural relations between the objects. Considering the complexity of the notation, we do not think that is very likely. We therefore propose to build guidance and support in the CSAV environment in such a way that (a) problem solving states are directly linked to objects that are in focus and (b) objects themselves signal to what sort of other objects they can or need to be connected and (c) categorical objects such as 'solution', 'type of cause', 'type of intervention' are available to represent a number of subsumed objects.

In the next section we review some of the demands that we formulated in chapter 2 and reconsider their translation to services offered in CSAV environments.

Global functional requirements for a CSAV environment

Here we present a first version of the global functional requirements for a computer-supported environment for collaborative solving of ill-structured problems by the use of external representations. The CSAV environment offers the users facilities to create external representations of argumentation when solving social science problems.

The general goal of the system is to offer support by helping to explicate the argumentation used in the stages of analysis, solution and evaluation in solving ill-structured problems in a graphical way. This goal is different from the goal of CSCL environments. In CSCL environments one tries to evoke argumentation (Veerman, 2000) in order to stimulate knowledge co-construction. In our situation two levels apply. On one level, the problem-solvers are arguing about elements of the analysis and the solutions on which they hold different views or opinions. We have included elements in the representational notation to support this type of argumentation, but the evidence in the literature and our own results do not indicate that much interaction can be expected here. Users at best will consolidate the results in the representation. On the other level we stimulate our users – even if they agree on all aspects of the analysis and solution – to provide the underlying reasoning, as if they were arguing with an external evaluator.

Although we did not go through the exercise, the idea behind these specifications is to ultimately develop a script (Dillenbourg, 2002) of the problem solving process not with the intention to define rigorously how the problem solvers have to operate, but to describe in a more precise way the tasks and processes. Note, that by incorporating this aspect, we have defined a number of shared representations that problem solvers need to use. Whether they actually can, is a different matter.

We will present the functional requirements per problem solving state identified in chapter 2. Before going in to these details, we present a number of additions to the representational notation presented in chapter 2. That notation was based on an eclectic blend of contributions from different research backgrounds. We do not have the illusion that the explorations reported in the previous chapters give us a research base on which we can decide what to *exclude* from the representational notation. We will therefore concentrate on the elements that seemed missing. First, as indicated by the EPA protocols, we need a way to express more abstract causes, such as 'social problems', or 'infrastructure'. We will refer to this category as 'cause type'. Second a similar category is needed for more abstract interventions. The category 'intervention strategy' was introduced to the coding scheme following the analyses reported in chapter 4. Note that these abstract type categories correspond to the way experts describe causes and solutions. Third, the category 'plan' is added to the representational notation so as to express a set of related interventions that is meant as a solution.

Table 6.1 was presented earlier in chapter 2 as a summary of the cognitive and communicative demands of the problem solving states of orientation, solving and evaluation. These demands were subsequently translated to the representation, communication and guidance and support services offered by CSAV environments. We will reconsider these demands and their translations to services given the results of the studies reported in the previous chapters.

Table 6.1
Cognitive and communicative demands of the different problem solving states

Problem solving states	Cognitive demands	Communicative demands
Orientation	Problem Representation Constraints Problem structuring Establish shared representations	Issue-based communication Brainstorm Build trust Establish common ground
Solving	Apply macro-operators to produce solutions Use topic and control representations Maintain coherence Maintain accuracy Maintain plausibility	Topic-based discussion Maintain common ground Maintain focus Conflict detection and resolution Knowledge negotiation
Evaluation	Evaluate solutions Evaluate constraints Evaluate process	Negotiate criteria

One of the first things to consider is whether in our studies we see signs that these problem solving states apply. From the literature we know that experts typically spend a considerable amount of time and effort in problem *orientation*. They try to explicate problem constraints, criteria and try to develop a first representation of the problem (Voss, Tyler, et al., 1983). In our protocols, students typically start with a short orientation, sometimes directly followed by ideas about solutions, in which they address some, but not all, of the orientation topics mentioned in the table. As expected, the protocols show that students spend considerable amounts of time on *solving*, that is formulating interventions that may eliminate the causes of the problem. As we saw earlier, the argumentation to support analysis and solution was often missing in the protocols. Finally, we saw little evidence of *evaluation*. Again, this is not very surprising: solutions more or less emerge.

Support of problem solving states

In this section we detail the functional requirements for a CSAV environment that supports collaborative solving of social science problems using external representations of argumentation. We will not deal with any of the administrative tasks that a system obviously needs to support, such as maintaining project data and status, as well as data on users or user groups.

In chapter 2 we identified a number of problem-solving states and the different cognitive and communicative demands of these states. These then were translated into the services offered by CSAV environments. The way that the system realizes this support is by offering different modes of collaboration, different representational notations and a number of functions that allow monitoring of developing solutions.

We will assume that the environment allows sharing, both in synchronous as well as in asynchronous mode of the external representations. One may consider other options however in which (a subset of) the graphical data are converted into forms readable by other applications. Compendium® (Selvin, 2003) or Mifflin® as it has been called as well, gives a good idea of the current possibilities. The system is based on open standards (Java, SQL, XML and others) and offers interfaces to various other tools and environments, such as the D3E discussion environment, Word, Excel, Visio and others. Compendium® is extensible, that is one can add to the behavior of nodes in the representation and Compendium supports meta-data (such as data on plausibility or agreement). The reader is referred to the Compendium website for information (http://www.compendiuminstitute.org) about recent versions of the tools.

We also assume that the system maintains an agenda with a list of tasks that the participants have to accomplish or objects they wish to include in the representation. Most of these tasks are put on the agenda by the participants. If, however, one of the guidance and support services notes that particular aspects of the representation need the attention of the problem solvers, it can take the initiative to formulate a topic for the agenda. The agenda acts as a control mechanism: the problem-solvers can decide on the basis of the topics on the agenda what they are going to address and in which mode (if that selection is available).

Support for orientation

In this phase initial problem representations are developed and finally issues for further exploration are identified. The style of communication in this phase is described as issue-based or conversational. In the asynchronous environment ACT® this phase is implemented as a discussion that has only a chronological order (Duffy et al., 1998; Sloffer et al., 1999). Our students, as described above, do not enter into lengthy problem orientation activities, but rather they are focusing on one aspect, such as the constraints, or their initial ideas about the causes of the problem. In chapter 2 we were rather concerned that in the orientation phase representational notations such as IBIS be used to prevent premature commitment to particular views on the problem. However, as described in the review of the studies reported in this thesis, we have found no substantial evidence in the protocols of any domain or stakeholder specific views on the problems. This brings us to the more fundamental question whether we need any other representational service in this state than one that allows inspections of created representations. This does not imply that the objects and relations contained in the representational notation cannot be discussed in this phase. It only means that there are no means to make *changes* to the external representation during this phase. What we are proposing here, is that students in an orientation state are stimulated to consider specific objects – ‘constraints’, ‘cause type’, ‘intervention type’ and maybe ‘data’. Once they decide to add any of these objects to the representation they can create such an item on the agenda. When they enter the solving state, the agenda items can be inserted into the representation. All this is only meant to stimulate that students consider more than one particular aspect of a problem. As reported in the previous chapters we found that our students often focused on particular aspects of the problem whilst ignoring others.

As the Ghent experiment indicated it is important that in this phase, as well as in others, the communication services of the CSAV ensure that the communication between the learners is *anchored* in the external representation they are co-constructing. Communication services in this state facilitate that learners express basically unordered issues or that they engage in structured brainstorm on constraints, or potential causes of the problem. More detailed ‘negotiation for meaning’, in which the participants establish initial common grounds on the issues that they have to tackle, has to be supported as well. The results of this phase are a list of issues (questions) to be addressed, and objects to be inserted into the external representation. The type of support is offered by implementing an interface with a tool like GroupSystems® that has all the facilities for electronic brainstorming, raising issues and knowledge negotiation.

Support for solving

For the solving state we stated demands such as 'apply macro-operators to produce solutions' and 'use topic and control representations'. Furthermore, demands where formulated for functionalities that maintain coherence, accuracy and plausibility in the proposed solutions. In the protocols of the beginning experts, we find no evidence of the use of topic and control representations. Obviously the EPA students who tried to apply techniques of Logical Framework Analysis (LFA) used both topic and control representations, but we also noticed that they did not succeed in maintaining coherence within and between the LFA representations they produced. Other topic and control representations (Alpay et al., 1998) have not been identified in the protocols. We identified a need to support shared and permanent representations, but the protocols contain no indications that warrant such a requirement.

Considering the studies reported in the previous chapters, there is one major issue in the way students approach these problems: they tend to focus on a subset of topics, ignoring others and they fail to address the structural relations between elements of the analysis and solution of the problem. As a starting point to offer support, we intend to use the guidance effect (Suthers, 2001) of the external representation to have learners address the structural aspects of the problem and its solution. External representations have a guidance effect on several levels. At the representational level they support to express particular information better than expressing other sort of information. One may expect that what is easier to express will more often be expressed. On another level, a representational notation provides the learners with a task structure: their interpretation of the task is such that they will use the objects and relations offered to them. Finally, the guidance effect manifests itself in the topics addressed in the dialogues between the learners. The question here is how this guidance effect can be explored. In the previous chapters we saw that students failed to use the structural relations between the objects in the representational notation. Expert teams on the other hand use external representations as data-structures, that is, as structurally related pieces of information (Alpay et al., 1998). Guiding or scaffolding learners to make adequate use of the representational services our CSAV offers cannot suffice by signaling the available objects and relations alone. Here we propose that the guidance operate on several levels:

First, the system adapts to the problem solving states that the learners are operating in. When the learners enter an orientation state, they can put issues on an issue list and they are limited to discussing only a particular type of content: constraints, causes (and cause types), interventions (and intervention strategy) and plan. In this state these objects can be created, but they cannot be incorporated in the external representation. Only when the learners enter a solving state they can work on all elements of the external representation. Newly defined objects can be taken from the agenda and incorporated into the representation.

Second, the objects entered into the external representation 'know' about other types of objects they can be connected to. Objects have *visible connectors* that signal to what type of object they can be related. Thus the objects afford that learners enter relations to particular types of objects. If the states of the connectors can be checked, the system can alert the learner that relations are missing and should be considered.

There are other functionalities that reach related goals. For example, a Collaboratory Notebook® note 'knows' which note types are legitimate follow-ups. IBIS-based graphical environments can enforce that the diagrams adhere to the grammar of the notation. Finally, one might even consider creating complete data-structures (with objects such as Cause, Data, Intervention). Whether learners perceive and use the support remains to be seen, however.

The system supports *separate* representation and discussion of constraints, features and risks. Constraints as defined here correspond to the subgoals defined in DRL. Evaluation of the solution involves a discussion of the claim that solution X satisfies the constraints defined here. Risks are collections of external threats to the solution, for example intervention X will only work, if financing is continued beyond the current schedule. Risks are related to interventions, but are collected for inspection so as to allow a risk assessment for the solution. Statements about constraints, features and risks are debatable claims.

In order to support the demands on the coherence, accuracy and plausibility of proposed solutions, guidance and support services are needed. We propose that the system collects ratings and computes aggregate data that can be used to evaluate the solution and its underlying argumentation. Here we mention only two services: agreement ratings and plausibility ratings. It is conceivable that other functions are supported, but they will require more advanced solutions. As an example, consider monitoring coherence. This requires that the system can reason about the relations and dependencies in an argumentation *structure*, for instance by applying truth maintenance mechanisms. Data on agreement and plausibility are easier to collect and process. Ratings on agreement are collected at the claims (objects and relations) in the representation. The system maintains agreement ratings on individual and group level thus allowing to make the differences between stakeholders visible. This corresponds to what is called a 'viewpoint' in the SIBYL® prototype (Lee, 1990). The system will signal disagreements than are larger than a threshold value.

The system will also collect users' ratings on the plausibility of claims entered. The system maintains plausibility ratings on individual and group level. We do not envisage a sophisticated plausibility management system as in SIBYL®; perhaps a simple filter mechanism will suffice. The most important function is that the system will signal when the variance in the ratings between individuals and / or groups is larger than a threshold value.

Users as well as the system can generate agenda items in this phase. The system will generate items where it spots lacking material (for instance no data to support the causal analysis), no criteria to evaluate solutions, no constraints posted) lack of agreement, or conflicts. The system can keep track of plausibility levels and report those that are below certain thresholds for individuals as well as groups.

Support for evaluation

Evaluation, as defined previously, answers the question as to whether the correct problem has been solved using the correct procedures and whether the constraints are met. In the protocols analyzed for the studies reported in the previous chapters, we find only a few segments that can be described as being concerned with evaluation. This does not mean that participants do not spend effort on evaluating solutions. In a process in which constraints get rephrased, and interventions get refined the problem solvers may be surprised to find that the intervention they agreed on meets the constraints and criteria defined during the problem solving process.

The other evaluation issues identified here are harder to answer: if the problem-solvers have not reached sufficient common ground on representations and operators, they will not agree on any solution. The only support that the system can offer is to pinpoint to areas in the representations where there is no overlap or insufficient agreement.

Here again we may consider whether we need representational services beside those that allow inspection of the representation. Again, we are inclined to believe that such is not the case. Finally, we assume that evaluation is a state that is best executed in a synchronous mode and we propose that GroupSystems® is a candidate environment to support this state.

This is only a first draft of a global functional requirements specification for a computer-supported environment that supports collaborative problem solving of ill-structured problems. It is an ambitious set of functional requirements for a potentially very advanced system if all the requirements are to be supported. According to the six stage model of Kirschner (2002) a next step would be to determine how learners perceive the support and subsequently whether they actually use the support. A more piecemeal engineering approach would address the core representational services first and gradually move to more advanced services. The core of the support that we intend learners to perceive and use is to not use particular objects of the representational notation in isolation, but to also consider the structural relations contained in the notation.

Directions for future research

The work reported here has brought together research from different areas to better understand how external representations of argumentation can support collaborative solving of a type of ill-structured problem called social science problems in the psychological tradition and 'wicked problems' in the planning and design tradition. The type of problems that we are interested in are referred to as 'ill-structured problems' or social science problems in the psychological literature. The research tradition in design and business problem solving has been more inspired by work on 'wicked problems'. The latter tradition has resulted in the first systems that supported argument visualization in problem solving. We built on the research tradition of external representation of argumentation in CSCL and planning and design, and incorporated insights from policy evaluation to design a first representational notation to be used in a CSAV environment.

In a review of CSAV environments rooted in these different traditions we identified communicative services, representational services and guidance and support services. We developed a model that identified the cognitive and communicative demands of the states in this type of problem solving and we linked these demands to the representational, communicative and guidance and support services of CSAV environments. In the previous section we have outlined a number of global functional requirements for a CSAV environment based on this model and on the results of the studies reported in the previous chapters. We can see two directions for future research, one to further test and develop the model brought forward here, the other to make progress in the development of a CSAV environment that offers the services described globally in the functional requirements.

There are several issues related to the model that need additional work. First, the model described how solving ill-structured problems involves multiple agents who all bring their views to the problem, that is different representations, operators and views on relevant constraints and criteria. Our approach, in fact stated as a demand that these multiple agents could express their multiple representations. In our studies, we could demonstrate different approaches in the mono-disciplinary teams of the beginning experts, but we failed to demonstrate any such differences in the multidisciplinary teams of the EPA students. Other research in the context of CSAV environments presented by Conklin, Selvin and Van Gelder in a recent volume on CSAV (Kirschner, Buckingham Shum and Carr, 2003) as well is all based on informal analyses only.

We need to understand under which conditions we can evoke expression of these different views on the problem. The settings that we used in our experiment may have been less than favorable in that respect. Other problems and other conditions, including individual preparation for the problem solving sessions, may be needed to stimulate these different views to be expressed.

In the functional requirements specified in the previous section of this chapter we described how we intend to use the guidance of the representational notation to stimulate that students pay attention to the structural aspects of the objects in the external representation. To study these and other ´affordances´ of external representations a series of well controlled experiments are necessary.

Our studies have concentrated on the content of the students dialogues only. As we indicated above, this was the correct choice as long as the development of a representational notation was concerned. Now, this perspective needs to be amended by one in which the mechanisms of collaboration can be better understood. Analysis of the communicative functions in the dialogue may also help to reveal how the services described in the functional requirements influence the collaboration between students.

We need better understanding of the cost of adding external representations to a problem solving context. As we have indicated elsewhere (Van Bruggen, Kirschner, & Jochems, 2002) there is no reason to assume that external representations will always result in cognitive off-loading. There are good reasons to assume that the reverse may occur as well. More work is needed to establish the trade-off between the affordances of the representational notation and its implementation on the one hand and problem-solving processes and outcomes on the other. This requires studies that are better controlled than the explorative studies we conducted here.

A second area for future research is more developmental in nature. As we described above, we have only touched upon the first stages of a development model for CSCL environments. To develop a CSAV environment that will support collaborative solving of social science problems many more of these steps need to be taken. We will outline a few of the next steps.

First, we have to address more in full the third stage of the development model of Kirschner (2002), that is work out the constraints. We have worked with small scale, synchronous problem solving settings. A more realistic setting for students of the Open University of the Netherlands however is that of an asynchronous setting, with facilities for occasional synchronous work, in which the students spend several weeks to solve similar problems. As we described above this may also be a setting in which different backgrounds of the students may get expressed in different viewpoints on a problem. Replication of some of the studies in hardware and software constellations in which students can work on problems for a sustained period is necessary. Here attention needs to be given to tasks and contexts in which different viewpoints on a problem can and will be expressed.

Second, a next step in the development of a CSAV is to investigate how students perceive and experience the support offered by external representations. Before implementing services in a computer-based environment, small scale, pen and paper experiments can be used to get a first idea of the perception and use of the representational notation and in particular the idea of guiding students towards working out relations with other objects.

Third, we have proposed some rather drastic limitations on the interaction with external representation by restricting editing of the external representation to the solving state of the problem solving process. In the orientation and evaluation state, learners can use the agenda to transfer objects for use in the external representation. This is an attempt to not only keep learners focused on activities, but also to encourage them to reflect on the current state of the problem solving process. Obviously, we need empirical results to check whether this corresponds to the perception and experience of the learners. Ultimately, their activities decide on any design we propose.

Summary

This thesis is concerned with the question how computer-based creation of external representations can support collaborative problem solving of social science problems (Voss, Tyler, & Yengo, 1983). An example of such a problem is that of how to reduce school drop-out. This type of problem is ill-structured and ill-defined: the goals, constraints and criteria for a solution are either not stated or are ambiguous. Characteristic to social science problems is 'delayed evaluation'. It may take years before the effects of interventions become visible. All this implies that solving this type of problem is to an important extent an argumentative process that already starts with the first analysis step: Is school drop-out an educational problem, or a social-economic problem as well? This argumentative process extents to the evaluation of proposed solutions. Since no one can prove that a solution will be effective, one can only provide arguments why it is likely to work. The external representations addressed in this thesis are informal or semi-formal graphical representations of the argumentation found in the analysis and solutions of the problem.

In chapter 2 the backgrounds of the research are sketched. A model of collaborative solving of social science problem is presented (based on Van Bruggen, Boshuizen & Kirschner, 2003). The model makes a distinction between orientation, solving and evaluation states in problem solving. For each state cognitive and communicative demands are formulated and these demands are subsequently related to functionalities (services) found in CSCL environments and environments aimed at supporting planning and design.

Most important to our research are the functionalities with which external representations can be created and the underlying representational notation, that is the objects and relations that can be used in the external representations. Chapter 2 presents a first version of this notation that incorporates concepts taken from the literature on evaluation research (Rossi, Freeman, & Lipsey, 1999).

Chapter 3 gives a sobering account of an attempt to incorporate elements of the representational notation in an existing environment for collaborative inquiry (Belvédère®). Second year Education students of the University of Ghent were randomly assigned to pairs and to either a Low or a High Detail condition. In the Low Detail instruction they were instructed to use no more than two objects in Belvédère® that were described as 'assumptions' and 'facts'. In the High Detail condition four objects were available that were described in more detail. The student pairs were instructed to compile recommendations on the swift implementation of inclusive education. They were presented with background material to support their analysis and recommendations. It was predicted that in the High Detail condition students would prepare richer and denser diagrams, that they would remember more of the background material and their recommendations. We also expected that in the High Detail condition the processing of the material would be harder and more time-consuming. The data however, all pointed to quite opposite directions.

One of the technical problems that plagued this experiment led to a condition in which students worked together sharing one computer. Exploration of the data, using this 'condition' indicated that the type of collaboration may be at least as important as the representational tools offered. A better controlled and balanced replication with additional data on the coordination processes between students is necessary to substantiate the current results.

Considering these sobering results, the attention was shifted toward further development and testing of the representational notation rather than developing and testing representational tools. Chapter 4 reports a first attempt to validate the representational notation developed. The notation was adapted to a coding scheme for dialogue analysis. Dialogues were coded from three mono-disciplinary teams of American students: Philosophers, Educators and Debaters that addressed the problem of how to reduce school drop-out. The results of these analyses were encouraging in that the coding scheme could not only cater for a substantial amount of the dialogues, but could be used to make visible the different approaches of the teams. All three groups addressed subsets of the analyses, leaving out other relevant aspects. One of the conclusions therefore is that process support is needed that helps the problem solvers to construe and maintain the coherence of the objects. Given the small empirical base of these observations and the limited reliability replication is needed.

In chapter 5 we report on the results of offering process support by incorporating the representational notation and the methods and techniques of Logical Framework Analysis. LFA prescribes a systematic procedure in which to work from problem analysis to the definition of a solution (project). In the study reported here four groups of four students participated who were enrolled in the Maastricht International Masters of European Public Affairs. These groups solved the school drop-out problem as well. Two groups were left free to select a method, two others were told to use LFA. Their dialogues were analyzed using the coding scheme. A remarkable result was that the protocols showed hardly any sign of the expertise of the students, who had varying educational backgrounds. The analysis indicated that LFA offered some process support, but on the other hand we found indications in the protocols that LFA does not encourage students to substantiate the underpinnings of their analysis or solution.

In chapter 6 we present first global functional requirements for an envisaged environment. The core of the recommendation here is that the environment should encourage students to construct and maintain the structural cohesion in the external representation. It is suggested that the system should adapt to the different problem solving states. We conclude with directions for future research and development.

Samenvatting

Dit proefschrift handelt over de vraag hoe computergebaseerde externe representaties het gezamenlijk oplossen van sociaal-wetenschappelijke problemen (social science problems in de terminologie van Voss, Tylor en Yengo, 1983) kunnen ondersteunen. Een voorbeeld van zo'n type probleem is de vraag hoe vroegtijdige schoolverlating kan worden teruggedrongen. Dit soort problemen zijn zwakgestructureerd en -gedefinieerd: de doelen, randvoorwaarden of criteria waaraan een oplossing moet voldoen, en/of de stappen die kunnen leiden tot een oplossing zijn niet gegeven of zijn ambigu omschreven. Kenmerkend voor sociaal-wetenschappelijke problemen is 'uitgestelde evaluatie': het kan jaren duren voordat het effect van maatregelen zichtbaar wordt. Dit alles brengt met zich mee dat het oplossen van dit soort problemen vooral een argumentatief proces is dat al begint met de analyse van het probleem - is vroegtijdige schoolverlating een onderwijsprobleem of (ook) een sociaal-economisch probleem? - en dat zich uitstrekt tot de beoordeling van alternatieven. Niemand kan bewijzen dat een oplossing werkt, men kan dat hooguit aannemelijk maken. De externe representaties waartoe wij ons beperken, zijn informele of semi-formele, grafische weergaven van de argumentatie die ten grondslag ligt aan de analyse en oplossing van het probleem.

In hoofdstuk 2 worden de achtergronden van het onderzoek beschreven. Er wordt een model gepresenteerd voor het gezamenlijk oplossen van sociaal-wetenschappelijke problemen (ontleend aan Van Bruggen, Boshuizen en Kirschner, 2003). Het model maakt een onderscheid tussen oriëntatie, oplossing en evaluatie stadium. Voor ieder stadium worden vereisten geformuleerd op cognitief en communicatief gebied. Deze vereisten worden vervolgens gerelateerd aan een aantal functionaliteiten die worden aangetroffen bij bestaande omgevingen voor computerondersteund samenwerkend leren (CSCL) en/of bij omgevingen die zijn ontwikkeld voor zakelijke toepassingen.
Het meest belangrijk voor ons onderzoek zijn, uiteraard, de functionaliteiten waarmee externe representaties kunnen worden vervaardigd en de daaraan ten grondslag liggende representationele notatie: de objecten en relaties waarvan gebruik kan worden gemaakt in de externe representatie. Gebruikmakend van concepten van ondermeer de literatuur op het gebied van evaluatieonderzoek (Rossi, Freeman en Lipsey, 1999) wordt in hoofdstuk 2 een eerste versie van een dergelijke notatie gepresenteerd.

Hoofdstuk 3 beschrijft de ontnuchterende ervaring die is opgedaan met het incorporeren van elementen van de representationele notatie in een bestaande omgeving voor gezamenlijk onderzoekend leren (Belvédère®). Tweedejaars Opvoedkunde studenten (N=95) van de Universiteit Gent werden random aan paren en aan een Hoog- resp. Laag-Detail conditie toegewezen. In de Laag-Detail conditie werden ze geïnstrueerd om gebruik te maken van slechts twee Belvédère-objecten die werden omschreven als 'assumpties', resp. 'feiten'. In de Hoog-Detail conditie waren vier objecten beschikbaar die in meer detail werden beschreven. De studenten werd gevraagd aanbevelingen op te stellen voor een snelle invoering van inclusief onderwijs. Ze konden daarvoor gebruik maken van achtergrondmateriaal over inclusief onderwijs. Voorspeld werd dat studenten in de Hoog-Detail conditie rijkere en hechter verbonden diagrammen zouden maken en dat ze meer van het achtergrondmateriaal en van hun eigen aanbevelingen zouden onthouden. Er werd verwacht dat in de Hoog-

Detail conditie de verwerking van het materiaal lastiger zou zijn en meer tijd zou vergen. De data echter bleken in volstrekte tegenspraak met deze voorspellingen.

Een van de technische problemen waar dit experiment door werd geplaagd had ons genoodzaakt studenten samen achter een computer te laten werken. Exploratie van de data, gebruikmakend van deze 'conditie' wees er op dat de wijze waarop studenten samen moeten werken een factor van belang kan zijn. Herhaling van het experiment onder beter gecontroleerde omstandigheden en met meer zicht op coördinatie handelingen van de studenten is nodig.

Op grond van de resultaten van dit onderzoek is de aandacht in het onderzoek meer komen te liggen op verdere ontwikkeling en uittesten van de representationele notatie dan op het ontwikkelen en testen van tools. In hoofdstuk 4 wordt verslag gedaan van een eerste poging om de ontwikkelde representationale notatie te valideren. De notatie werd omgewerkt tot een coderingsschema voor het analyseren van dialogen. Dialogen werden geanalyseerd van drie verschillende mono-disciplinare groepen van Amerikaanse studenten: filosofen, onderwijskundigen en 'debatteurs' die zich bezig hielden met de vraag hoe vroegtijdige schoolverlating kon worden tegengegaan. De resultaten van de analyse waren in die zin bemoedigend dat het coderingsschema niet alleen een substantieel deel van de dialogen kon 'vangen', maar ook kon met het schema de verschillende benaderingen van de groepen zichtbaar worden gemaakt. Alle drie de groepen bleken zich te beperken tot een deel van de analyse, waarbij ze andere elementen systematisch buiten beschouwing lieten. Een van de conclusies van dit onderzoek is dat meer ondersteuning nodig is die de probleemoplossers de samenhang van de objecten laat construeren en bewaren. Daarnaast is, gezien de geringe empirische bases en de niet te hoge betrouwbaarheid, herhaling van het onderzoek nodig.

In hoofdstuk 5 werd onderzocht of de procesondersteuning kon worden gerealiseerd door incorporatie van de representationele notatie in de methoden en technieken van Logical Framework Analysis (LFA). LFA volgt systematisch een aantal stappen van een probleemanalyse naar een definitie van een oplossing (project), waarbij gebruik wordt gemaakt van externe representaties om de analyse van het probleem en de werking van de oplossing zichtbaar te maken. In dit onderzoek participeerden vier groepen van vier studenten die in Maastricht een Internationaal Masters programma volgden over Europees beleid (European Public Affairs) en die waren getraind in de beginselen van LFA. Ook deze groepen kregen het probleem van de vroegtijdige schoolverlating voorgeschoteld. Twee groepen werden vrijgelaten in de te volgen methode; twee groepen werden geïnstrueerd gebruik te maken van LFA. Hun dialogen werden geanalyseerd met het coderingsschema. Opvallend was de onzichtbaarheid van de specifieke expertise van de deelnemers die diverse studieachtergronden hadden. De analyse van de dialogen laat zien dat toepassing van LFA wel enige procesondersteuning lijkt te bieden. Anderzijds vonden we ook aanwijzingen in de protocollen die er op wijzen dat LFA onvoldoende stimuleert tot onderbouwing van de analyse zowel als de oplossing van het probleem.

In hoofdstuk 6 presenteren we eerste globale functionele eisen aan een omgeving zoals beoogd. De kern van die aanbevelingen komt er op neer dat moet worden bevorderd dat studenten de structurele samenhang in de externe representatie realiseren en vasthouden. Daartoe wordt tevens voorgesteld om het systeem zich aan te laten passen aan de verschillende stadia in het probleemoplosproces. We sluiten af met aanbevelingen voor verder onderzoek en ontwikkeling.

References

Alpay, L., Giboin, A., & Dieng, R. (1998). Accidentology: an example of problem solving by multiple agents with multiple representations. In M. W. van Someren, P. Reimann, H. P. A. Boshuizen, & T. de Jong (Eds.). *Learning with multiple representations* (pp. 152-174). Amsterdam: Pergamon.

Barron, B. (2000). Achieving coordination in collaborative problem-solving groups. *Journal of the Learning Sciences, 9*, 403-36.

Bell, P. (1997). Using argument representations to make thinking visible for individuals and groups. In R. Hall, N. Miyake, & N. Enyedy (Eds.). *Proceedings of CSCL '97: the second international conference on computer support for collaborative learning* (pp. 10-19). Toronto: University of Toronto Press.

Bell, P., (2001). Using argument map representations to make thinking visible for individuals and groups. In T. Koschmann, R. Hall, & N. Miyake (Eds.). *CSCL2: carrying forward the conversation* (pp. 449-485). Mawah, NJ: Erlbaum.

Bell, P., Davis, E. A., & Linn, M. C. (1995). The Knowledge Integration Environment: theory and design. In J. L. Schnase & E. L. Cunnius (Eds.). *Proceedings of CSCL '95 : the first international conference on computer support for collaborative learning* (pp. 14-21). Mahwah, NJ: Erlbaum.

Boshuizen, H. P. A., & Tabachneck-Schijf, H. J. M. (1998). Problem solving with multiple representations by multiple and single agents: an analysis of the issues involved. In M. W. van Someren, P. Reimann, H. P. A. Boshuizen, & T. de Jong (Eds.). *Learning with multiple representations* (pp. 137-151). Amsterdam: Pergamon.

Bromme, R., Nückles, M., & Rambow, R. (1999). Adaptivity and anticipation in expert-laypeople communication. In S. E. Brennan, A. Giboin, & D. Traum (Eds.). *Psychological models of communication in collaborative systems. Papers from the 1999 fall symposion* (pp. 17-24). Menlo Park, CA: AAAI. Available from http://www.aaai.org

Buckingham Shum, S. (2003). The roots of computer supported argument visualization. In P. A. Kirschner, S. J. Buckingham Shum, & C. S. Carr (Eds.). *Visualizing argumentation: software tools for collaborative and educational sense-making* (pp. 3-24). London: Springer.

Buckingham Shum, S. J., MacLean, A., Bellotti, V. M. E., & Hammond, N. V. (1997). Graphical argumentation and design cognition. *Human-Computer Interaction, 12*, 267-300.

Buckingham Shum, S., & Hammond, N. (1994). Argumentation-based design rationale: what use at what cost? *International Journal of Human-Computer Studies, 40*, 603-652.

Carroll, J. M., & Moran, T. P. (1991). Introduction to this special issue on design rationale. *Human-Computer Interaction, 6*, 197-200.

Chen, H.-C. (1990). *Theory-driven evaluations*. Newbury Park, CA: Sage Publications.

Chryssafidou, E. (1999, August). *Computer-supported formulation of argumentation: a dialectical approach.* Paper presented at the symposium 'Belvedere: review and new applications', August 29, 1999, Heerlen.

Conklin, E. J., & Begeman, M. L. (1988). gIBIS: A hypertext tool for the exploratory policy discussion. *ACM Transactions on Office Information Systems, 6*, 303-331.

Conklin, E. J., & Weil, W. (1997). *Wicked problems: naming the pain in organizations*. Retrieved June 05, 2001, from http://www.gdss.com/wp/wicked.htm

De Jong, T., Ainsworth, S., Dobson, M., Van der Hulst, A., Levonen, J., Reimann, P., Sime, J.-A., van Someren, M. W. ; Spada, H., & Swaak, J, (1998). Acquiring knowledge in science and mathematics: the use of multiple representations in technology-based learning environments. In M. W. van Someren, P. Reimann, H. P. A. Boshuizen, & T. de Jong (Eds.). *Learning with multiple representations* (pp. 9-40). Amsterdam: Pergamon.

Deutsche Gesellschaft für Technische Zusammenarbeit (GTZ). (1997). *Ziel Orientierte Project Planung - ZOPP; eine orientierung für die planung bei neuen und laufenden projekten und programmen*. [Goal Oriented Project Planning – GOPP; an orientation for planning new and on-going projects and programs]. Darmstadt, Germany: GTZ.

Dillenbourg, P. (2002). Over-scripting CSCL: The risks of blending collaborative learning with instructional design. In P. A. Kirschner (Ed.). *Three worlds of CSCL: Can we support CSCL* (pp. 61-91). Heerlen, The Netherlands: Open University of the Netherlands.

Duffy, T. M., & Cunningham, D. J. (1996). Constructivism: implications for the design and delivery of instruction. In D. H. Jonassen (Ed.). *Handbook of research for educational communications and technology* (pp. 170-198). New York: Macmillan Library Reference.

Duffy, T. M., Dueber, B., & Hawley, C. (1998, November). *Critical thinking in a distributed environment: a pedagogical base for the design of conferencing systems* (CRLT Technical report No 5-98). Bloomington, IN: Indiana University, Center for Research on Learning and Technology.

Edelson, D., & O'Neill, D. K. (1994). *The CoVis Collaboratory Notebook: computer support for scientific inquiry*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans.

Edelson, D. C., O'Neill, D. K., Gomez, L. M., & D'Amico, L. (1995). A design for effective support of inquiry and collaboration. In J. L. Schnase & E. L. Cunnius (Eds.). *Proceedings of CSCL '95: the first international conference on computer support for collaborative learning* (pp. 107-111). Mahwah, NJ: Erlbaum.

Eggers, H. W. (1998, May/June). Project cycle management revisited. *The Courier,* 69-72.

Erkens, G. (1997). *Coöperatief probleemoplossen met computers in het onderwijs; het modelleren van coöperatieve dialogen voor de ontwikkeling van intelligente onderwijssystemen* [Cooperative problem-solving with computers in education; modelling cooperative dialogues for the development of intelligent tutoring systems]. Unpublished PhD dissertation, University of Utrecht, The Netherlands.

European Commission. (2001, March). *Manual project cycle management*. Brussels: EuropAID Co-operation Office.

Gasper, D. (1999, January/February). Problems in the logical framework approach and challenges for 'project cycle management'. *The Courier,* 75-86.

Gasper, D. (2000). Evaluating the 'logical framework approach' -- towards learning-oriented development evaluation. *Public administration and Development: an international journal of training, research, and practice, 20*(1), 17-28.

Goel, V., & Pirolli, P. (1992). The structure of design problem spaces. *Cognitive Science, 16*, 395-429.

Herring, S. (1999). Interactional coherence in CMC. *Journal of Computer-Mediated Communication*, 4. Retrieved July 28, 2000 from http://jcmc.huji.ac.il/vol4/issue4/herring.htm

Hewitt, J., Scardamalia, M., & Webb, J. (1997). *Situative design issues for interactive learning environments: the problem of group coherence*. Paper presented at the Annual Meeting of the American Educational Association, Chicago. Retrieved July 28, 2000 from http://twilight.oise.utoronto.ca/abstracts/situ_design/

Kanselaar, G., Erkens, G., Andriessen, J., Prangsma, M., Veerman, A., & Jaspers, J. (2003). Designing argumentation tools for collaborative learning. In P. A. Kirschner, S. J.

Buckingham Shum, & C. S. Carr (Eds.). *Visualizing argumentation: software tools for collaborative and educational sense-making* (pp. 51-73). London: Springer.
Kirschner, P. A. (2002). Can we support CSCL? Educational, social and technological affordances of learning. In P. A. Kirschner (Ed.). *Three worlds of CSCL. Can we support CSCL*? (pp. 7-47). Heerlen, The Netherlands: Open University of the Netherlands.
Kirschner, P. A., Buckingham Shum, S. J., & Carr, C. S. (Eds.). (2003). *Visualizing argumentation: software tools for collaborative and educational sense-making*. London: Springer.
Kirschner, P. A., van Bruggen, J. M., & Duffy, T. 2003). Validating a representational notation for collaborative problem solving.In Wasson, B., Ludvigsen, S., & Hoppe, U. (Eds.). *Designing for change in networked learning environments* (pp. 163 - 172). Dordrecht: Kluwer Academic Press.
Kolodner, J., & Guzdial, M. (1996). Effects *with* and *of* CSCL: tracking learning in a new paradigm. In T. D. Koschmann (Ed.). *CSCL, theory and practice of an emerging paradigm* (pp. 307-320). Mahwah, NJ: Erlbaum.
Kunz, W., & Rittel, H. (1970, July, reprinted May 1979). *Issues as elements of information systems* (Working Paper No. 131). Center for Planning and Development Research: University of California at Berkeley. Retrieved August 6, 2002, from http://www-iurd.ced.berkeley.edu/pub/WP-131.pdf
Lee, J. (1990). SIBYL: a qualitative decision management system. In P. H. Winston & S. A. Shellard (Eds.). *Artificial Intelligence at MIT; expanding horizons* (pp. 105-133). Cambridge, MA: MIT Press.
Lee, J., & Lai, K.-Y. (1991). What's in design rationale? *Human-Computer Interaction, 6*, 251-280.
Miao, Y., Holst, S., Holmer, T., Fleschutz, J., & Zentel, P. (2000). *An activity-oriented approach to visually structured knowledge representation for problem-based learning in virtual learning environments*. Paper presented at the fourth international conference on the design of cooperative systems (COOP 2000), Sophia Antipolis, France. Retrieved May 26, 2000 from ftp://ftp.darmstadt.gmd.de/pub/concert/publications/camera_ready_coop2000.pdf.
Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.
Norman, D. (1998). *The invisible computer: Why good products can fail, the personal computer is so complex, and information appliances are the answer*. Cambridge, MA: The MIT Press.
Paolucci, M., Suthers, D., & Weiner, A. (1995). *Belvedere: stimulating students' critical discussion*. CHI95 conference companion, interactive posters, May 7-11, Denver CO, 123-124.
Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. London: Sage Publications.
Reitman, W. (1965). *Cognition and thought*. New York: Wiley.
Rittel, H. W. J. (1984). Second-generation design methods (interview by Donald P. Grant and Jean-Pierre Protzen). In N. Cross (Ed.). *Developments in design methodology* (pp. 317-328). Chichester: John Wiley & Sons. (Originally published in The DMG 5th anniversary Report: DMG Occasional paper No. 1 (1972), pp 5-10)
Rittel, H. W. J., & Webber, M. M. (1984). Planning problems are wicked problems. In N. Cross (Ed.). *Developments in design methodology* (pp. 135-144). Chichester: John Wiley & Sons. (published earlier as part of 'Dilemmas in a general theory of planning', *Policy Sciences,4*, 1973, 155-169)
Rossi, P. H., Freeman, H. E., & Lipsey, M. W. (1999). *Evaluation: a systematic approach* (6th ed.). Thousand Oaks, CA: Sage Publications.

Sartorius, R. (1996). The third generation logical framework approach: dynamic management for agricultural research in projects. *European Journal of Agricultural Education and Extension, 2*, 49-62.

Scardamalia, M. A., & Bereiter, C. (1994). Computer support for knowledge building communities. *The Journal of the Learning Sciences, 3*, 265-283.

Scardamalia, M., Bereiter, C., Maclean, R. S., Swallow, J., & Woodruff, E. (1989). Computer-supported intentional learning environments. *Journal of Educational Computing Research, 5*, 51-68.

Selvin, A. M. (2003). Fostering collective intelligence: helping groups use visualized argumentation. In P. A. Kirschner, S. J. Buckingham Shum, & C. S. Carr (Eds.). *Visualizing argumentation: software tools for collaborative and educational sense-making* (pp. 137-163). London: Springer.

Sloffer, S. J., Dueber, B., & Duffy, T. M. (1999, February). *Using asynchronous conferencing to promote critical thinking: two implementations in higher education* (CRLT Technical Report No. 8-99). Bloomington, IN: Indiana University, Center for Research on Learning and Technology.

Stahl, G. (2001). WebGuide: guiding collaborative learning on the web with perspectives. *Journal of Interactive Media in Educatio.* Retrieved August 06, 2002, from: http://www-jime.open.ac.uk/2001/1

Stenning, K. (1998). Representation and conceptualisation in educational communication. In M. W. Van Someren, P. Reimann, H. P. A. Boshuizen, & T. de Jong (Eds.). *Learning with multiple representations* (pp. 320-333). Amsterdam: Pergamon.

Stenning, K., & Oberlander, J. (1995). A cognitive theory of graphical and linguistic reasoning: logic and implementation. *Cognitive Science, 19*, 97-140.

Suthers, D. (1995). *Designing for internal vs external discourse in groupware for developing critical discussion skills.* CHI' 95 research symposium, Denver.

Suthers, D. (2001). Towards a systematic study of representational guidance for collaborative learning discourse. *Journal of Universal Computer Sciences, 7*(3), 254-277.

Suthers, D. (in press). Studies of Representational Support for Collaborative Inquiry with Belvedere. In J. Andriessen, M. Baker, & D. Suthers (Eds.). *Confronting cognitions: arguing to learn*. Dordrecht: Kluwer Academic Press.

Suthers, D., & Weiner, A. (1995). *Groupware for developing critical discussion skills.* Retrieved July 14, 1998, from http:/www-cls95.indiana.edu/cscl95/suthers.html

Suthers, D., Weiner, A., Connelly, J., & Paolucci, M. (1995). *Belvedere: engaging students in critical discussion of science and public policy issues*. Paper presented at the 7th World conference on artificial intelligence in education (AI-ED 95), Washington.

Suthers, D. D. (1999). Effects of alternate representations of evidential relations on collaborative learning discourse. In C. Hoadley & J. Roschelle (Eds.). *Computer support for collaborative Learning; designing new media for a new millenium: collaborative technology for learning, education, and training. CSCL 99, December 12-15, Palo Alto, CA* (pp. 611-620). Palo Alto, CA: Stanford University.

Suthers, D. D., Toth, E. E., & Weiner, A. (1997). An integrated approach to implementing collaborative inquiry in the classroom. In R. Hall, N. Miyake, & N. Enyedy (Eds.). *Proceedings of CSCL '97: the second international conference on computer support for collaborative learning* (pp. 272-279). Toronto: University of Toronto Press.

Team Technologies Inc (1999). TeamUp-PCM [Computer software].

Toulmin, S. E. (1958). *The uses of argument*. Cambridge, UK: Cambridge University Press.

Van Bruggen, J. M., Boshuizen, H. P. A., & Kirschner, P. A. (2003). A cognitive framework for cooperative problem solving with argument visualization. In P. A. Kirschner, S. J. Buckingham Shum, & C. S. Carr (Eds.). *Visualizing argumentation: software tools for collaborative and educational sense-making* (pp. 25-47). London: Springer.

Van Bruggen, J. M., & Dekeyser, H. (2001, August 30). Ontology and specificity for argumentation in Belvédère. In C. Pontecorvo (Chair), *Argumentation and reasoning in different discursive and learning environments*. Symposium conducted at EARLI 2001, Fribourg.

Van Bruggen, J. M., & Kirschner, P. A. (in press). Designing external representations to support solving wicked problems. In J. Andriessen, M. Baker, & D. Suthers (Eds.). *Confronting cognitions: arguing to learn*. Dordrecht, The Netherlands: Kluwer Academic Press.

Van Bruggen, J. M., Kirschner, P. A., & Jochems, W. (2002). External representations of argumentation in CSCL and the management of cognitive load. *Learning and Instruction, 12*, 121-138.

Van Gelder, T. (2003). Enhancing deliberation through computer supported argument visualization. In P. A. Kirschner, S. J. Buckingham Shum, & C. S. Carr (Eds.). *Visualizing argumentation: software tools for collaborative and educational sense-making* (pp. 97-115). London: Springer.

Van Someren, M. W., Reimann, P., Boshuizen, H. P. A., & de Jong, T. (Eds.). (1998). *Learning with multiple representations*. Amsterdam: Pergamon.

Veerman, A. (2000). *Computer-supported collaborative learning through argumentation*. Unpublished PhD dissertation, University of Utrecht, The Netherlands.

Visser, W. (1990). More or less following a plan during design: opportunistic deviations in specification. *International Journal of Man-Machine Studies, 33*, 247-278.

Voss, J. F. (1991). Informal reasoning and international relations. In J. F. Voss & D. N. Perkins (Eds.). *Informal reasoning and education* (pp. 37-58). Hillsdale, NJ: Erlbaum.

Voss, J. F., Blais, J., Means, M. L., & Greene, T. R. (1986). Informal reasoning and subject matter knowledge in the solving of economics problems by naive and novice individuals. *Cognition and Instruction, 3*, 269-302.

Voss, J. F., Greene, T. R., Post, T. A., & Penner, C. (1983). Problem-solving skill in the social sciences. In G. H. Bower (Ed.), *The psychology of learning and motivation: Vol. 17. Advances in research and theory* (pp. 165-213). New York: Academic Press.

Voss, J. F., Perkins, D. N., & Segal, J. W. (Eds.). (1991). *Informal reasoning and education*. Hillsdale, NJ: Erlbaum.

Voss, J. F., Tyler, S. W., & Yengo, L. A. (1983). Individual differences in the solving of social science problems. In R. F. Dillon & R. R. Schmeck (Eds.). *Individual differences in cognition* (Vol. 1, pp. 205-232). New York: Academic Press.

Voss, J. F., Wiley, J., & Sandak, R. (1999). Reasoning in the construction of argumentative texts. In J. Andriessen & P. Coirier (Eds.). *Foundations of argumentative text processing* (pp. 29-41). Amsterdam: University of Amsterdam Press.

Wessner, M., Pfister, H.-R., & Miao, Y. (1999). Using learning protocols to structure computer-supported cooperative learning. *Proceedings of the ED-MEDIA'99 - World Conference on Educational Multimedia, Hypermedia & Telecommunications*, Seattle, WA, June 19-24, 1999, pp. 471-476.

Zolin, R., Fruchter, R., & Levitt, R. E. (2002). *Simulating the process of trust: using simulation to test and explore a social process.* Retrieved August 10, 2002 from http://www.casos.ece.cmu.edu/conference2000/pdf/Roxanne-Zolin.pdf

Appendix The coding scheme used to analyze the protocols

ID	Code	Definition	Example
1	**Analysis**	Statements related to the (causal) analysis of the drop-out phenomenon.	
1.1	**Problem scope**	Statements that address how the problem is defined or delineated, or which part(s) of the problem will be addressed.	Yeah, secondary schools and does that include high school? I don't think that very many people, I mean drop-out of elementary school
1.2	**Hypothesis**	Statements that explicitly mention (candidate) cause(s) or explanation(s) for drop-out.	Low SES. Um, and that could be families and communities.
1.3	**Data**	Statements that contain explicit mentioning of evidence related to one or more hypotheses	Yeah um one of the main ones that that worked was just create a more positive um student um whatchamacallit um help create a positive thing within students what's it called
1.3.1	Experiential data	Evidence based on personal experience, observation or induction from experience or observations	Well, it's true like I have a really good friend who was this woman named Sheryl and she was she dropped out and she was cool as heck she was really awesome person but she just thought that school was a big waste of her time.
1.3.2	Authority data	Evidence based on publications, empirical data, scholars	… there's uh studies from James Kohlmer who's been working in schools since 1968 and all the schools that go that go under this principle of involving …
1.4	**Principle** (warrant)	Statements that contain explicit mentioning of the process dynamics that relate data and drop-out.	there's a ton of research um out there um that that states that with the with more parental involvement more positive parental involvement there's a increase in student achievement and also um it minimizes drop-out rates
1.5	**Research question**	Statements that phrase (a) question(s) for further research. Note: participants were instructed to compile a list of research questions to be answered before their presentation. Only statements that pose a research question to be put on this list are coded here.	Has there been anything done about charter schools that aid in dropping the reducing the drop-out rate or... we should look at as if we're going to if we're going to even think in terms of incentives to finish school, um, that's something we should research for the next meeting..
2	**Constraints & Features**	Statements that formulate (sub)goals that the recommended actions should satisfy or constraints that the recommended actions have to meet	

ID	Code	Definition	Example
2.1	Features	Statements that express desirable properties of the recommendation. Note that this code refers to features of recommendations that are of a general nature.	One thing I think is important and might help us out too is that if we if we tend to focus on trying to um maybe affect variables that can be changed instead of saying you know we've got this plan that's gonna you know do something remarkable I mean, maybe we can operate on the assumption that we're gonna do what we can with what is there right now
2.2	Conditions and Risks	Statements that express (pre)conditions to the recommendations or identify risk factors.	I wonder if we have to work within like resources. We should ask that question I wonder if we're allowed that? does our solution have to be like an all or nothing, like can they take parts of the solution that they liked?
3	**Intervention logic**	Statements relating to the concrete actions to be taken (interventions), their expected immediate results (outcomes), intervening mechanisms (assumptions) and their expected long-term effects (results).	
3.1	**Interventions**	Statements about concrete actions (including critical remarks; questions about et cetera)	I mean, I do think that there's something to be said though for incorporating into the solutions something that some kind of a program or an idea or a suggestion that does something to try an make school better… I mean, because I think you can try an address drugs an you can try an address teen pregnancy and all of those things, but those are long term … maybe if we have a combination of solutions …
3.2	**Outcomes**	Statements that contain descriptions of immediate outcomes of interventions (including critical remarks; questions about et cetera)	there's certain external factors that you just you just can't control for you know there's you can't control necessarily for the population there and the SES and some other factors but you certainly can control for some other internal factors you know like you're saying— teacher training, staff— we need maybe to think about some things we can do to

ID	Code	Definition	Example
3.3	**Assumptions**	Statements about intervening mechanisms that link immediate outcomes to results. Note: here and elsewhere the statement can also mean questioning or challenging the relation	… the way that we try to deal with these problems like the drug problem for example is just scream and holler about like, you know, don't use drugs don't use drugs which I think is probably good thing to do, but all of those things become so like moralistic for some high school kid who's involved with that I don't know if those things are very effective. But that might be one of the things that we can offer um, the governors is the suggestion that uh one way to improve retention is to expand the number of alternative high schools other than college prep high schools.
3.4	**Effects**	Statements containing descriptions of the effects of the intervention and the way these effects are established (operational definitions, et cetera).	guess so so what you're saying is that one easy obvious way uh, to like improve the school retention rate is to raise the age at which you of how long you have to stay in school.
4	**Present solution**	Statements on how to present results	
4.1	**Goals**	Statements about desired features of the presentation	
4.2	**Plan and outline**	Statements about general content and content ordering of the presentation	No, I think with that what we what we could do is make up an RFP for the for the governors. We'd say, if you want to read it, here it is. Um, maybe 4-5 pages long and that could include research and stuff.
4.3	**Content**	Statements about concrete content of presentation	
5	**Other**	Remaining segments go here	

Curriculum Vitae

Jan van Bruggen (1953) completed a Gymnasium education and then studied Education at the University of Amsterdam. During and following his studies he worked at the Center for Research in Higher Education (COWO) where he was engaged in research and development in the area of study skills and reading strategies. Together with Marcel Mirande he developed and evaluated courses in 'Schematiseren' – a study technique that learned students to represent the concept structures in study texts. At COWO he worked with the Plato system for computer-based training; a highly innovative environment that initiated many of the innovations that now (20 years later) are found on every desktop computer.

From 1981 to 1984 he co-pioneered a small firm that specialized computer-based training and distance teaching material and was engaged in the development of courses for various banks and insurance companies and the training and supervision of in-company teachers and educational technologists.

From 1984 to 1987 he returned to the University of Amsterdam for a project on control strategies during reading. After completing the project and after three reorganizations at the University he accepted the offer to join Courseware Europe, then the market leader in computer-based training. He participated in the Esprit project Eurohelp and was involved in the proposal for the Delta project ACES an became coordinator of that project.

In 1991 he joined the Open University of the Netherlands and was involved in several externally funded projects such as the EU projects EAST and CTA and one of the TACIS projects. He had the responsibility for planning and control of several of these externally funded projects.

In 1993 he initiated a project that developed a knowledge-based system capable of representing OUNL curricula and the associated, complex regulations. The system was demonstrated to handle more than 95% of the examination cases flawlessly. In 1996 and 1997 the OUNL mapped the history of its students onto its redefined curricula. The prototype was revamped to do the initial processing of some 10,000 student records.

He was involved in various OUNL projects when he started to participate in the Otec research program. Next to the research reported in this thesis his interests are in the area of computer-based essay scoring, in particular the use of Latent Semantic Analysis.