Use of expertise by beginning experts solving wicked problems

Running head: USE OF EXPERTISE IN SOLVING WICKED PROBLEMS

Use of Expertise by Beginning Experts Solving Wicked Problems

Jan M. van Bruggen and Paul A. Kirschner*

Open University of the Netherlands

*and University of Utrecht

Th. M. Duffy

Indiana University

Chad S. Carr

Express, Limited Brands

Wim M.G. Jochems

Open University of the Netherlands

Abstract

In this article we study how groups of beginning experts solve a type of problem that has been referred to as ‘wicked’ or ‘social science problem’. An example of such a problem is that of school drop-out, which can be defined as a school and curriculum problem, a socio-economic problem, a cultural problem, a behavioral problem, et cetera. Mono-disciplinary teams of beginning experts in different fields addressed the problem of advising the State Board of Governors on measures to reduce school drop-out. One group demonstrated expert-like behavior in solving the problem. The other groups concentrated on particular aspects of the problem and its proposed solution whilst neglecting other aspects. Their approach contained a mixture of (beginning) expert and novice behavior, demonstrating that the expert-novice continuum is not clear-cut for these types of problem. The use of external representations that signal which aspects of the problem solving process need additional effort is discussed as a promising support, in particular to inter-disciplinary teams.

Use of Expertise by Beginning Experts Solving Wicked Problems

The kind of problem dealt with here has been referred to as ‘wicked’ in the literature on planning and design (Kunz & Rittel, 1970). Such problems are often typified in psychological literature as ill-structured and more specific as ‘social science problems’ (Voss, Greene, Post, & Penner, 1983; Voss, Tyler, & Yengo, 1983). Solving this type of problem is elusive because it often is not clear what area of expertise is (most) relevant to solving it. Consider the example problem reported on in this article: What can you advise the Board of Governors to help reduce high school drop-out? Drop-out itself can be approached as an educational problem, but it also makes sense to consider it a social problem, or an economic problem, and so forth. These approaches are related to criteria by which interventions to reduce drop-out may be evaluated. Moreover, other criteria may have to be considered as well, such as political feasibility and social and moral acceptability.

Voss studied this type of problem, which he called ‘social science problems’, using expert-novice paradigms (Voss, Green et al., 1983; Voss, Tyler et al., 1983) and his classic example problem, is that of how to increase the crop production in the Soviet Union. Voss, Tyler et al. (op. cit., p. 208) 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)”. They found that a typical expert solution consisted of a few abstract solutions for a general cause, whereas novices tried to isolate different causes and defined solutions in terms of eliminating these individual causes.

Another characteristic of such social science problems is that the problem solvers cannot be certain beforehand whether a chosen solution will work because it may take years before a solution is implemented and produces any observable results. Voss, Tyler 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 may work, or is more likely to work than other solutions 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.

In the literature on planning and design these problems have been called 'wicked problems' (Kunz & Rittel, 1970; Rittel & Webber, 1984). According to Conklin and Weil (1997) such problems (a) are composed of an interlocking set of issues and constraints, rather than a definitive statement of the problem itself; (b) have 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; (c) have constraints on the solution which change over time; and (d) often have no definitive problem, so that there is also no definitive (or right) solution; problem solving ends when time, money, energy, et cetera runs out; and (e) solutions cannot be tested and revised: one cannot try-out a highway trajectory, to use the classic example of Rittel.

A major difference between the two strands of research is that whereas traditional problem solving research, such as the expert-novice paradigms used by Voss, has used single agent settings, those concentrating on 'wicked problems' have emphasized the involvement of multiple agents, or stakeholders, each of whom may bring a particular viewpoint on the problem.

More recent research has studied the way in which *teams* of problem solvers coordinate their problem solving activities by sharing data and/or problem solving operators (Boshuizen & Schijf, 1998). Alpay, Giboin and Dieng (1998), for example, studied the use of representations by multi-disciplinary teams of engineers and psychologists who analyzed traffic accidents. They found that the teams used permanent representations that corresponded

to systems, procedures and models on a regular basis in standard situations. Temporary representations were built dynamically during the analysis of a specific accident. Some representations were shared. For instance, all engineers and psychologists shared a *simple* functional model of the driver, but only the psychologists used a richer version of this model. Finally, control representations, such as models and phase decompositions, were used to guide operations on topic representations.

The ultimate aim of our research is to support multi-disciplinary teams that are trying to solve social science problems by providing them with the types of (external) representations Alpay et al. (1998) identified. As a preliminary to that, we studied how mono-disciplinary teams from different backgrounds approached a social science problem, namely that of high school drop-out. The background was chosen so as to compare expertise in the content of the problem to that in the process of solving the problem. Any solution to a social science problem requires formulation of a coherent set of (a) identified causes and (b) interventions of which one can argue convincingly that they will remove these causes, whilst (c) operating within acceptable constraints. We selected three groups to bring specific expertise to this problem solving. Educators were supposed to bring *content knowledge* of the educational system and educational policies. The two other groups were expected to bring *process knowledge* to the tasks: Philosophers were expected to have expertise in the areas of systematic analysis, reasoning and argumentation and ethical issues, and Debaters were expected to bring in process expertise in systematic discussion of problems and (persuasive) arguing for specific solutions and presenting these in a convincing way to their audience.

Methods

Participants

The participants in this study were American university students with different backgrounds (Debating, Philosophy and Education). Data were collected on a total of six

groups, each consisting of three participants with the same background. All participants, with the exception of the Debaters, were graduate students. Data were collected into sets, each containing three groups and they became available for analysis separately. Initial results for the first data set were reported elsewhere (Kirschner, Van Bruggen, & Duffy, 2003). Here the results for all data are presented in more detail. Because of the differences in the length of the protocols analyzed and the changes to the protocol coding scheme used, we will present the results of the two sets separately.

The participants were confronted with the following problem: The annual meeting of the Board of State Governors has asked a team of experts to prepare recommendations to reduce high 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 15 minute presentation for the Board.

Scoring

An analysis scheme to code the content of the typed written protocols of the first 90-minute sessions (see the Appendix) was constructed to capture social science problem solving. The scheme has four main categories (referred to in capitals throughout this article) and several subcategories. The main category *Analysis* was used for statements relating to the scope of the problem, the causes of the problem (hypotheses) and evidence (data) brought forward in support of these hypotheses. *Constraints & Features* was used to code for statements about desirable or necessary features of solutions as well as preconditions and risks. *Intervention Logic* was used for statements dealing with concrete interventions to remove perceived causes; statements about the expected effectiveness of interventions, such as short-term and long-term outcomes and statements about how the interventions would lead

to these effects. *Presentation* was used for statements relating to the preparation of the presentation that the subjects, according to their instruction, had to deliver.

In the coding scheme all statements relating to a topic, including critical remarks or questions, are scored. This means that data and not the operators (‘state’, ‘question’, ‘elaborate’ et cetera) are considered. After the analysis of the first set small improvements in the details of the coding scheme were implemented. The data reported here are corrected for these differences. The full coding schemes are available from the first author.

The coding was based on the typed protocols only (no audio or video data were used). All protocols were coded by two trained coders who were unaware of the backgrounds of the participants. Reliabilities were calculated using the data from the two coders. The data reported here, are based on the coding of one coder for each protocol. The typed protocols were segmented on the basis of the speech-turns taken in the dialogue. Since a turn may address different topics, coders were instructed to code for the main idea expressed in the turn. In addition, they were asked 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.

The coders classified the segments of the first data set in one of four main categories (*Analysis, Constraints & Features, Intervention Logic, Presentation*). For the first data set there were agreements on the main categories of 72% (Philosophers=P), 77% (Educators=E) and 79% (Debaters=D) protocols. The Cohen’s Kappas for these protocols were .59, .66 and .69 respectively. For the second set there were agreements on the main categories of 81% (P) 87% (E) and 84% (D). The Kappas here were .71, .81, and .77 respectively.

In this article we will concentrate on the main categories and a small set of subcategories, dealing with identified *causes* of drop-out, the proposed interventions (concrete actions to be taken) and *constraints* that solutions had to fit in. In order to get an

impression of the breadth of the content covered in these subcategories we had one coder typify the contents of statements that were coded as hypotheses, constraints or interventions. The coder was instructed to use, wherever possible, the descriptions found in the protocols themselves, such as “curriculum-related causes”, or “parent-child-relation” to typify contents.

Results

The number of segments in the protocols of the two data sets is reported in Table 1. As this table shows, the length of the protocols (in segments) varies between the data sets (DS) as well as between the groups. The protocols in DS1 contain, overall, fewer segments than those in the second data set. Within this set there are substantial differences between the groups: the Debaters’ protocol contains almost four times the number of segments than the Philosophers. This is partly due to the length of the Philosophers’ speech-turns (= segments) which was almost twice the length (in words) of the segments of the other groups. The number of segments in DS2 shows less variance, although the Debaters’ protocol contains almost twice the number of segments as that of the Educators.

Overall, the protocols in DS2 contain more segments than DS1, possibly because the groups in the second set all use (more than) the full length of the time available (90 minutes) for the task. Considering these differences, the two data sets are analysed and reported separately.

***

INSERT TABLE 1 ABOUT HERE

***

Data Set 1

An impression of the relative number of segments dealing with the topics represented by the main categories in the coding protocol is represented in Table 2. All groups concentrate on *Intervention Logic*, which is not surprising considering the instruction they

had been given. The category *Analysis* scored second. This coding category is used for statements about causes of the problem, data relevant to causes, and the scope of the problem. Important differences are visible in the category *Constraints & Features*, where segments are coded that deal with necessary or desirable conditions for any solution. The Philosophers concentrate more on this category than the other two groups do.

We use Chi-square tests on the distribution of codes in the protocols of the groups. Strictly speaking, the data do not meet all the assumptions of a Chi-square test, because one cannot assume that the codes within a protocol are independent. Rather than refraining from statistical analysis, we use a conservative significance level of 1% and interpret the results of the test with some caution. This being set, a Chi-square test of the distribution shows deviations from the expected values that are statistically significant ($X^2(6) = 153.61$, $p < .001$). The standardized residuals in the table, which are distributed as *z*-scores, indicate which cell frequencies deviate significantly from the expected values. Using a 1% significance level standardized residuals are significant only if their absolute value exceeds 2.58. These residuals indicate that the Philosophers concentrated less than expected on *Analysis* and more than expected on *Constraints & Features*. The Educators concentrated less on *Intervention Logic* than expected and more than expected on *Presentation*, while Debaters, finally, addressed *Constraints & Features* as well as *Presentation* less than expected.

***

INSERT TABLE 2 ABOUT HERE

***

The data in Table 2 tell us how often the groups deal with topics in a category. These data do not tell us, for example, whether a particular cause of drop-out is discussed in depth, or whether several causes are considered. For a more detailed analysis we decided to concentrate on the dialogue dealing with the core content of a solution: the identified causes,

the proposed concrete interventions (actions) and the constraints (requirements any solution should meet). We are interested in the *breadth* of the discussion in the groups – how many different causes, interventions and constraints are considered – as well as the *depth* of the discussion, how often these causes, interventions and constraints are discussed. The number of segments dealing with each of the different causes, interventions and constraints is used as a measure of the *depth* of coverage.

Figures 1 through 6 present an attempt to visually represent the depth and the breadth simultaneously. In the figures the x-axis represents breadth of the discussion. In Figure 1, for example, we see that the Philosophers formulated two different causes, whereas the Debaters formulated ten. The y-axis, with *cumulative* frequencies of the segments dealing with the different causes identified, represents *breadth*. Cumulative frequencies are used here to correct for differences in absolute numbers between the groups and to make visible how the groups divide their attention between the different types of causes identified. Consider the following example: The Philosophers mentioned two types of causes and discussed these in four segments. In the Debaters' protocol ten different causes were identified in the 67 fragments that were dealing with causes. In order to make the data of the groups more comparable, the causes were ordered in descending order of frequency and *cumulative* frequencies were then calculated and plotted. If the same depth of treatment were given to all different causes, the resulting graph would be a straight line starting at the point 1/n, where n is the number of different causes. Most of the graphs, however, show an initial steep slope, which indicates that most of the discussion is devoted to a few topics. The flattening of the curve indicates that other topics were discussed in a few fragments only. Thus, Figure 1 shows that the Philosophers concentrated on one single cause and the other groups concentrated on two causes in particular, accounting for 50% of the segments.

***

INSERT FIGURE 1 ABOUT HERE

***

The cumulative frequencies of the constraints identified are reported in Figure 2. The groups *appear* to use different approaches here, but this should be taken with a grain of salt. The Debaters produced 11 constraints in 8 protocol segments. As the straightening line in the graph indicates, most of these constraints were mentioned only once. The same is true for the Educators, who formulated 8 constraints in 9 segments. The graph shows that only the first constraint is mentioned twice. The Philosophers formulated 11 constraints in 23 segments. Five of these constraints are mentioned more than once.

***

INSERT FIGURE 2 ABOUT HERE

***

In contrast to the constraints, the interventions reported in Figure 3 are based on a substantial number of codes for all three groups. The data in Figure 3 show that the Debaters formulated five types of intervention and all of these (in particular the first three) were discussed several times. This points to a certain depth of coverage. In contrast, the Educators and the Philosophers formulated more different types of Interventions, but they clearly focused on only three or four.

***

INSERT FIGURE 3 ABOUT HERE

***

Some of the results presented here are what may have been expected. For instance, we expected that the Philosophers would not have any deep knowledge of the drop-out problem or measures to counter it, but that they would use their knowledge of argumentation and ethical aspects. Indeed, they more or less ignored the analysis of the problem and its solution,

and concentrated instead on the constraints: what type of intervention is acceptable; what are the standards that must be maintained? We expected that the Debaters would bring in process expertise in systematic discussion of problems and (persuasive) arguing for specific solutions. The Debaters mentioned several potential causes and interventions and engaged in extensive discussion about them. In doing so they generated the longest protocol and the largest number of segments, if only by frequently interrupting each other. We expected that the Educators would demonstrate content knowledge on the topic of drop-out and measures to counter it. As it turned out the group had a graduate student on board who studied the topic of drop-out. Especially through his contributions, the group did not spend great effort in analyzing the problem, its constraints nor its potential solutions. Much as Voss' expert they came up with a abstract cause and solution within minutes. This brings about the unfortunate position that expert problem solving is being demonstrated by not showing itself in the tables. This issue will be discussed in the following section.

Qualitative analysis

Due to the small sample size as well as following the tradition begun by Voss and his colleagues, the quantitative analysis is augmented here by a more qualitative analysis of the results. A qualitative inspection of the protocols corroborates the inferences drawn from the quantitative analysis of segments. The discussion of the Philosophers starts with delineating the problem: they discuss the meaning of high school 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 preparatory schools. In segment 19 and further, 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 (see Table 1 and Figure 2). Their next step is to define targets for intervention: high 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].

The *Philosophers* structure the solution space in terms of constraints, 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 this point they have considered incentive approaches in some detail, 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 by brainstorming on the possible causes (drugs, gangs, pregnancy, jobs, homelessness, et cetera) adding some experiential data to support the claims. 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. Their counts on analysis, in particular hypotheses, and interventions (see Tables 1 and 2) are produced by enumerating (and repeating) the list of causes, and discussing 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. Interestingly enough, they often fail to formulate a particular recommendation. 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 by considering whether the recommendations should target only those learners who can be helped directly by making changes to the school [segment 529-533], possibly even writing of other learners [segment 590]. The Debaters then come up with several school related recommendations: teacher quality, resource improvement, more extra-curricular activities, and so further.

It is interesting to note how both the Philosophers and the Debaters run into problems caused by their initial approach. The Philosophers do not 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 cause that turn out to be at variance with the social nature of the stated causes. Debaters also lost track of the larger goal of reducing drop-outs and began to focus on how to implement (or whether to implement because of social considerations) the interventions. For example, the implementation of support for pregnant girls lost track of how large of a cause of the drop out problem this was or even if the solution addressed the drop-out problem at all. Instead, they focused on the practical and social issues in implementing a reduction of teenage pregnancy.

The *Educators* obviously have more background knowledge to substantiate their analyses and recommendations. 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 such as low social economic status, families and communities, and motivation, but already in segment 21 the core of their analysis and recommendation is formulated, namely community and parental involvement. Authority data is presented that relates involvement to reduce drop-out. The participant who is working on a PhD in a closely related area, 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 the PhD student gives all sorts of information on charter schools, while another participant tries to formulate the points that would be appealing to the governors [segments 71-126].

In segment 131 the Educators start planning their presentation and anticipate reactions of the Board, including the issue of standardization [127-219]. This leads to a number of innovations they would like to include in their recommendations (e.g., 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 where 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 a type of 'confirmation bias' in the Educators protocol in that no alternative causes or interventions receive substantial attention and the participant who questions whether the recommendations will effectively reduce high school drop-out, is generally 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 Board of Governors.

The analysis thus far indicates that all groups more or less capitalize on their specific expertise, which leads them to concentrate on particular aspects of problem and the solutions, while ignoring others.

Second data set

The following results are based on DS2. The counts and standardized residuals for the main categories are reported in Table 3. A Chi-square test indicates significant deviations from expected values ($X^2(6) = 74.569$, $p < .001$).

For the Philosophers the results correspond with those from DS1. Here as well they concentrate on *Constraints & Features* and use fewer fragments on *Analysis,* including causes, than expected. For the Educators the results are different from DS1. There, they concentrated primarily on preparing a presentation and less than expected on Intervention Logic. Here, in DS2, they concentrate more on *Constraints & Features*. In both data sets the Debaters spent less time on the *Constraints & Features* than was expected and in both sets, in particular in the second, they concentrate their efforts on Analysis, while more or less ignoring *Constraints & Features*.

***

INSERT TABLE 3 ABOUT HERE

***

We analyzed the depth and breadth of the discussion of causes, constraints and interventions in DS2. Figure 4 summarizes the results for the discussion on causes. As far as

the breadth of coverage is concerned there are no discernible differences between the groups. All three formulate six types of hypotheses. The depth of the discussion is slightly different. The Debaters spent twice as many segments on causes than the other groups and they do so more evenly than the other groups, who tend to concentrate on two causes.

***

INSERT FIGURE 4 ABOUT HERE

***

The depth and breadth of the discussion of constraints show some remarkable differences between the groups. Figure 5 shows that the Educators and Debaters almost completely ignore constraints. In contrast, the Philosophers in both data sets discuss several constraints at great length. In DS2, they not only spent more segments (45) on this topic than the other groups, they also mention more different types of constraints. Finally, as the curve in the graph shows, they did not concentrate on a particular constraint but addressed several ones more than once.

***

INSERT FIGURE 5 ABOUT HERE

***

The depth and breadth of interventions is depicted in Figure 6. As the figure shows, the protocols of Philosophers and the Debaters are quite similar in the breadth and depth of their discussion of interventions. The Educators, in contrast, mentioned five interventions, but in their discussion concentrated on two interventions.

***

INSERT FIGURE 6 ABOUT HERE

***

We will not repeat the qualitative analysis of the protocols for DS2, but only add three short observations.

First, the Philosophers in both data-sets operated in a similar way. Although the Philosophers in DS2 demonstrated more knowledge about Educational policies, in the end, they as well concentrated on issues related to the acceptability of solutions, that is, they as well focus on constraints.

Second, the Debaters in DS2, although not showing the mechanical approach of their counterparts in DS1, worked with a similar collection of unrelated causes (such as drugs, security, learning disabilities) to which they tied one or more interventions, whilst ignoring constraints almost completely.

Finally, there appears to be an important difference between the Educator groups. In DS1, the Educator-team contained a PhD student who was familiar with the drop-out problem and who formulated their central recommendation (stimulate the creation of charter schools) within eight minutes. Most of the turns on hypotheses and data and constraints were exchanges in which this drop-out expert shared his knowledge on causes and research findings, as well as on stated Governors' policies. This Educator-group quickly analyzed the problem and then jumped to the presentation issues, rather than analyzing other potential causes and interventions besides the school-based ones. In the Educator-team in DS2, there was no expertise on the drop-out problem. In the protocol, the team returns time and again to several potential interventions without reaching firm conclusions on causes or interventions. The team hardly mentions constraints to the solutions.

Discussion

According to Voss, Tyler et al. (1983), experts describe the causes of the problem in more abstract terms and come up with more general, abstract solutions to which lower level problems are subordinate. Novices decompose a problem in low-level sub-problems to which

solutions are proposed. They, typically, fail to evaluate solutions in terms of constraints and do not specify sub-problems that may be encountered when proposed solutions are being implemented. Elsewhere (Kirschner et al., 2003; Van Bruggen, 2003) we presented a qualitative analysis of the first data set and described how each group tried to exploit its particular expertise, focusing on particular aspects of the problem, whilst ignoring others. The current analysis corroborates most of these findings.

The Philosophers approached the problem more or less as an ethical one and probed the acceptability of several, mostly theoretical, interventions. One could say that they, thus, tried to delineate the problem and the constraints for the solution. Using their expertise, the Philosophers did not dwell on possible causes, but concentrated on the acceptability of potential solutions (i.e., on the constraints), without coming up with feasible solutions. The Philosophers may have been novices to the domain of school drop-out, they are beginning experts in the area of the ethics related to interventions, and this is the problem aspect they focus on.

In both data sets the Debaters tried to work systematically from a set of proposed causes. The Debaters in DS1, in particular, first compiled a list of causes that they subsequently addressed one by one, trying to describe an intervention to deal with the effects of each cause (and get entangled in details of the implementation). Both Debater teams attacked the problem in a systematic way, showing the kind of process expertise that one can assume that they would have. They analyzed the problem and came up with potential solutions, without considering the constraints (as did the Philosophers) or the validity of their analysis (as did the Educators, see below). The Debaters seem to follow the ‘Identify and Eliminate’ strategy described by Voss, but their approach also mimics the novice tactic in decomposing the problem in unrelated causes that they seek to eliminate one by one and in failing to formulate and check on constraints.

In both data sets the Debaters and Philosophers appeared to follow the same approach, but the results for the Educators are less clear-cut. In DS1, the Educators had an expert PhD student on board who dominated the debate and made the team operate much like an expert would. They decomposed the problem into two causal factors, community and parental involvement, for which they added research evidence. They formulated constraints that were to-the-point and pragmatic and so tested their proposed solutions against policies that they knew the Board of Governors had already adopted. Their solution is strictly educational, although this was not a stated constraint, and when they realized that other factors may be important as well, they explicitly decided to restrict the scope of their solutions to the educational arena. In contrast, the Educators in DS2 clearly lacked such specific knowledge of the problem of high school drop-out and the policies of the Governors. Their protocol shows the group struggling to identify and structure causes of drop-out. They consider advising the Board of Governors to spend more money on research to establish the causes and risk factors. Being uncertain about causes, their interventions and final recommendations were, in their own words, not very convincing.

Taking together, these results indicate that all groups try to use whatever expertise they have to solve the problem of high school drop-out, but this also makes them focus on particular aspects of the problem and its solution whilst ignoring others.

Before presenting our conclusions, a number of limitations associated with the approach taken need be addressed. First, the focus was on the content of the dialogues and was modeled using the concepts of the representational notation. To be more specific, the objects, and not the relations, of the representational notation were used, and as a consequence the structure of the argumentation of the students cannot be modeled. Furthermore, only the individual segments were analyzed and the larger structures such as episodes devoted to a particular topic were ignored. The way in which the dialogue

progresses through a number of topics and how focus is maintained are not modeled in the coding scheme used. However, the analysis of the content topics, gives an accurate estimate of the focus for particular types of content.

Second, there are several important aspects of the dialogue that are not captured, such as the means (i.e., speech acts) by which students express their ideas and arguments, how they coordinate their actions, and how they negotiate knowledge. It is expected that the use of an external representation will eventually help students maintain focus and coordinate their actions, but this is not modeled in the coding scheme used.

Conclusion

The type of problem studied here can be defined according to various viewpoints. Although all groups eventually approached the high school drop-out problem as being a school-related problem, this was a clear decision since they were aware that other approaches also made sense. Only the Educators in DS1 were able to combine the problem definition, the identification of abstract causes, relevant constraints, and interventions which they could backup with data. In doing so, they mimic the behavior of Voss' experts. The other groups do not demonstrate typical novice behavior, as reported by Voss'. Rather, they exploit their own specific expertise which makes them act like experts on particular aspects of the problem solving process. The Philosophers in both data-sets did not even bother to come up with effective interventions. They systematically focused on *Constraints & Features*, in particular the demands that any intervention would have to meet. In both data sets the Debaters seem to follow the expert strategy of Identify and Eliminate causes: the Debaters work through a list of causes and formulate interventions for each of them. Yet, in both data-sets these causes are unrelated and the Debaters fail to consider constraints or test the effectiveness of the proposed interventions.

Our results indicate that for the type of problem studied here an expert-novice

distinction may be problematic. Our groups were composed of beginning experts in relevant domains and, except for the group with the drop-out expert, all groups demonstrated beginning expert behavior in *some* aspects of the problem solving process, while operating as novices in others. These results may be interpreted as offering a chance of improving the problem solving performance of beginning experts by scaffolding or coercing them to address all aspects of the problem solving process. External graphical representations may be used to signal what has been addressed and what awaits additional effort. Whether this would have helped our teams to achieve a better analysis and solution seems questionable, however. The teams still lack the detailed knowledge of real experts. A more promising approach, in our view, is to use these representations in multi-disciplinary teams as ‘partially shared’ representations of a full solution process, with experts contributing in more detail to the various parts.

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, The Netherlands: Pergamon.

Boshuizen, H. P. A., & 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, The Netherlands: Pergamon.

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

Kirschner, P. A., Van Bruggen, J. M., & Duffy, T. (2003). Validating a representational notation for collaborative problem solving. In B. Wasson, S. Ludvigsen, & U. Hoppe (Eds.). *Designing for change in networked learning environments* (pp. 163-172). Dordrecht, The Netherlands: Kluwer Academic Press.

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 on June 05, 2001, from http://www-iurd.ced.berkeley.edu/pub/WP-131.pdf

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).

Van Bruggen, J. M. (2003). Explorations in graphical argumentation. Unpublished PhD thesis. Open Universiteit Nederland, Heerlen, The Netherlands.

Voss, J. F., Greene, T. R., Post, T. A., & Penner, C. (1983a). 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., Tyler, S. W., & Yengo, L. A. (1983b). 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.

Appendix: Coding Scheme

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 what 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 it as 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	
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?

ID	Code	Definition	Example
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 and 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.
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).	I guess 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 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	

Footnotes

The analyses of the protocols were carried out with the MEPA program, made available by Dr. Gijsbert Erkens of the University of Utrecht.

Table 1

Number of segments in the protocols of Philosophers, Educators and Debaters in the data sets

	Group		
Data Set	Philosophers	Educators	Debaters
1	250	452	921
2	1115	740	1426
Total	1365	1192	2347
Coded	696	585	1230
	(51%)	(49%)	(52%)

Table 2

Counts and standardized residuals for the main categories in the first data set.

		Group		
Category		Philosophers	Educators	Debaters
Analysis	Count	31	88	210
	Std. Res.	-3.1	.0	1.6
Constraints & Features	Count	37	15	13
	Std. Res.	8.2	-.6	-4.0
Intervention Logic	Count	54	68	230
	Std. Res.	-.4	-2.7	2.1
Presentation	Count	14	54	26
	Std. Res.	-.3	5.7	-3.8

Table 3

Counts and standardized residuals for the main categories in the second data set.

		Group		
		Philosophers	Educators	Debaters
Analysis	Count	199	135	386
	Std. Res.	-2.4	-1.6	3.1
Constraints & Features	Count	45	4	13
	Std. Res.	5.5	-2.6	-2.9
Intervention Logic	Count	295	213	360
	Std. Res.	.7	1.9	-1.9
Presentation	Count	10	11	10
	Std. Res.	.0	1.7	-1.1

Figure Captions

Figure 1. Cumulative frequencies of hypotheses in data set 1 for Philosophers (P), Educators (E) and Debaters (D)

Figure 2. Cumulative frequencies of constraints in data set 1 for Philosophers (P), Educators (E) and Debaters (D)

Figure 3. Cumulative frequencies of interventions in data set 1 for Philosophers (P), Educators (E) and Debaters (D)

Figure 4. Cumulative frequencies of hypotheses in data set 2 for Philosophers (P), Educators (E) and Debaters (D)

Figure 5 Cumulative frequencies of constraints in data set 2 for Philosophers (P), Educators (E) and Debaters (D)

Figure 6. Cumulative frequencies of interventions in data set 2 for Philosophers (P), Educators (E) and Debaters (D)

Figure 1

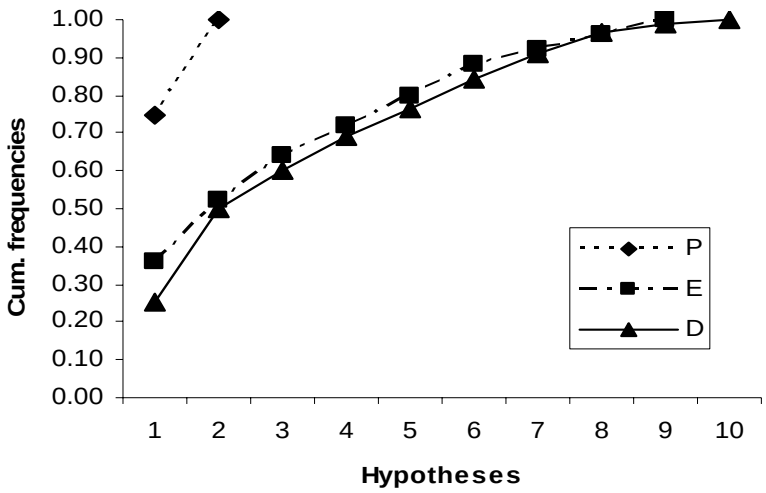

1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Cum. frequencies
1
2
3
4
5
6
7
8
9
10
Hypotheses
P
E
D

Figure 2

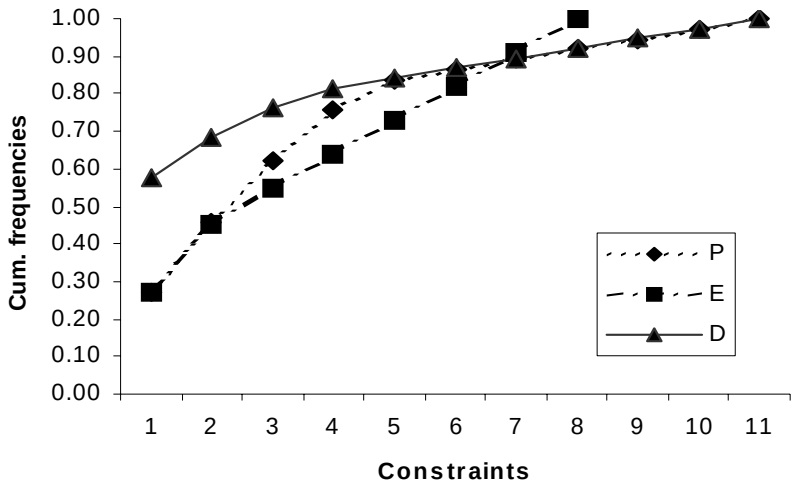

1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Cum. frequencies
1
2
3
4
5
6
7
8
9
10
11
Constraints
P
E
D

Figure 3

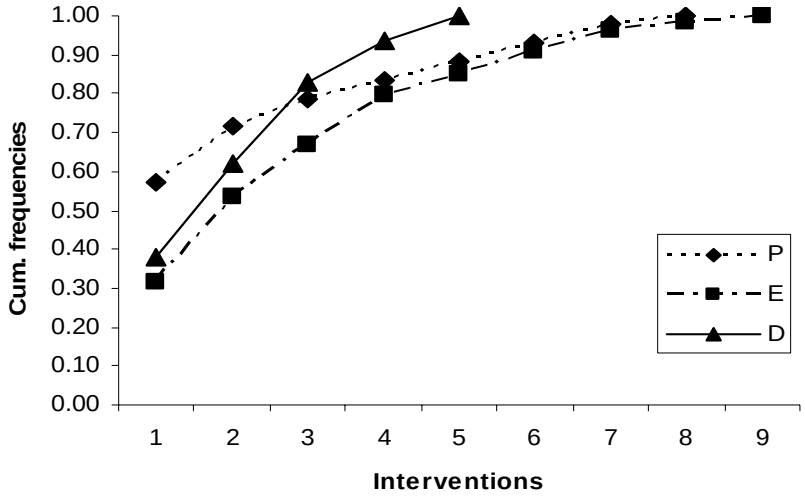

1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Cum. frequencies
1
2
3
4
5
6
7
8
9
Interventions
P
E
D

Figure 4

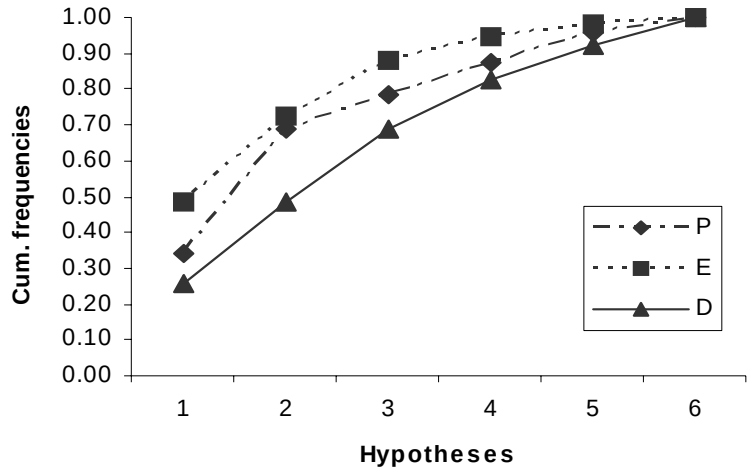

1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Cum. frequencies
1
2
3
4
5
6
Hypotheses
P
E
D

Figure 5

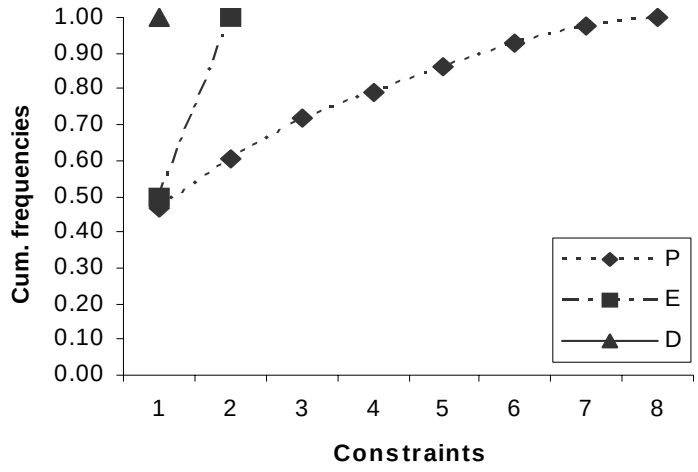
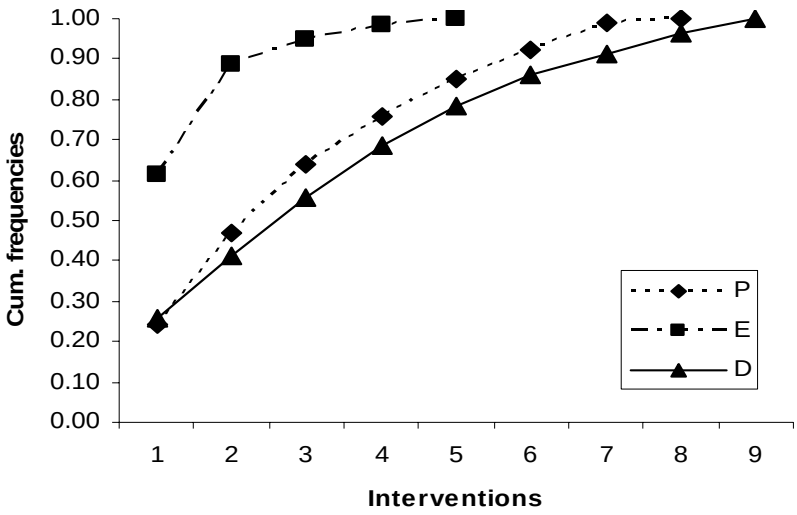

1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Cum. frequencies
1
2
3
4
5
6
7
8
Constraints
P
E
D

Figure 6


1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Cum. frequencies
1
2
3
4
5
6
7
8
9
Interventions
P
E
D