This is a pre-print version of the paper published in the **Studies in Educational Evaluation Journal, Vol 33 (2), pp. 175-196, June 2007**.

COMPUTER AIDED EVALUATION OF HIGHER EDUCATION TUTORS' PERFORMANCE

Michalis Xenos[1,*] and Thanos Papadopoulos[2]

[1]Hellenic Open University, 23 Saxtouri Str., GR 26223, Patras, Greece.
[2]Operational Research and Information Systems Group, Information Systems Research Unit, Warwick Business School, The University of Warwick, CV4 7AL Coventry, UK.

Abstract
This paper presents a method for computer aided tutor evaluation. The method is based on Bayesian Networks used for organizing the collected data about tutors and for enabling accurate estimations and predictions about future tutor behavior. The model provides indications about each tutor's strengths and weaknesses, which enables the evaluator to exploit strengths to the benefit of the University and offer advice for tutors' improvement. It also allows the evaluator to make hypotheses about potential tutor approaches and test the effect of such approaches on the educational procedure in advance. The paper briefly discusses Bayesian Networks and introduces a model that has been used in the Hellenic Open University for aiding tutor evaluation.

Keywords: Computer-aided evaluation, Higher education, Tutor evaluation, Bayesian Networks

Introduction

Over the last years tutor evaluation has been recognized as a necessity for many universities. Especially in the case of universities that employ a large number of external associates who have to renew their tutoring contracts with the institution every year, such as the case of the Hellenic Open University (HOU), the evaluation of tutors may be imperative. However, despite its necessity, tutor evaluation remains a difficult task to accomplish because of numerous crucial elements that need to be taken under consideration for an objective evaluation. In general, evaluating personnel is a procedure that is held "in a complex social context" (Scriven, 1995).

Methods used for measuring tutor performance aim at providing to the university that uses them an in depth understanding of its academic personnel. All academic institutions seek to improve the performance of their tutors, identify successful educational practices so that they are repeated and instruct tutors how to improve their teaching methods, better support and communicate with their students. An in depth understanding of the tutors' skills and abilities is important as it enables universities to place 'the right person to the right place'. For instance in the HOU case, an institution that offers distance learning, tutor assessment is crucial for ensuring quality of education, as tutors have to play multiple roles in the educational process: educational, consultative and supportive.

[*] Corresponding author, Tel.: +30 2610 367405; fax: +30 2610 367520
E-mail address: xenos@eap.gr

This paper introduces a method that can be used for the evaluation of tutors and a model which is used as a tool in the current HOU evaluation process of tutors. The presented method is generic and can be applied by any university-level institution. The presented model has been built specifically for the HOU; it has been used for tutors' evaluation during the academic years 2003-2004 and 2004-2005 and is still used in the current year. The necessity for modeling the process of tutor evaluation emanates from the fact that traditional assessment methods are, in most cases, 'memory' –as well as paper– based and classify tutors into categories based on their academic performance. The presented model takes into account the basic qualities that a tutor should have as well as his/her individual approach to issues related to his/her duties, e.g. the organization of an educational program. The presented model is a means for convenient data organization, while, using easy to understand rules of probability theory, it analyzes the possibilities that certain future educational actions occur, tests possible 'scenarios', predicts and estimates the corresponding results. Additionally, it provides an efficient way of identifying the reasons behind a successful educational action taken by a tutor during the academic year so that it can be repeated or of explaining why the results of an action met only a small percentage of the tutor's initial targets and expectations so that corrective action can be taken in the following academic years. A successful educational action may have to do e.g. with a tutor's initiative to communicate more effectively with his/her students. However, the effectiveness in communication can be depended on several factors/reasons, e.g. the way it is conducted (face-to-face interaction, e-mail, web, phone discussions, and Advisory Module Meetings between the tutors and the students). Hence, the success of such an initiative is contingent on the success of each of the aforementioned sub-factors. Finally, the model is self-learning as it uses its own results, either estimations or identifications of reasons that led to a given 'current' state, as feedback that can be used as input in future applications.

To apply the presented method, one would need previous years' data that will be used for building and 'training' the model as well as knowledge of the Bayesian Networks (BN) theory. However, after the initial model has been built, then the coordinator (or the body responsible for the tutors' evaluation) can use this model to feed in the current data without needing to know about the BN theory. In this manner, the model aids in keeping a record of tutors' data and uses such data for tutor evaluation. In addition to a posteriori evaluation, the model can also be used for testing future scenarios a tutor can follow in order to improve his/her performance in a specific field and assessing the impact that a specific improvement will have to the final evaluation outcome. For instance, if he/she needs to be improved in the way of supporting his/her students, then –because of the fact that providing support to students may be dependent on factors such as being patient, friendly and encouraging– he/she may need to be improved in some or in all of these factors; thus the designer of the model may test each one or all of these scenarios in order to get the outcome the improvement of the factors will have in shaping the student support factor and consequently the final estimation of the model for the tutor's academic performance.

Current higher education evaluation programs do not seek to judge, but to support teachers by introducing them to new methods and turning them into "reflective practitioners" (Gibbs and Coffey, 2000). For instance, there exist methods that either use self-concepts of teachers (Villa and Calvete, 2001), or measure the teachers' alternative methods of teaching (Coffey and Gibbs, 2002), or introduce "a way of scoring their instructors' performance" (Teerajarmorn *et al.*, 2003). The presented model introduces innovative features in the evaluation procedure and works complementary to already existing ones by combining the knowledge and the experience of an experienced evaluator with the advantages of an automated mathematical model.

The following sections introduce the model that has been implemented and used in the evaluation of the HOU tutors. After a short presentation of the HOU case and a brief introduction to Bayesian Networks, i.e. the mathematical basis of the evaluation model, the design and the application of the model are presented, basic concepts and fundamental issues are analyzed and specific examples of the model's application are given. Finally, this paper ends with conclusions that should be taken into consideration by those who intend to design and implement a similar model in the future.

Presentation of the particular case

The presented study is based on data collected in the HOU. The HOU started to operate in 2000 and offers educational programs leading to a bachelor's degree based on the Open and Distance Learning model. It has adopted a modular system, which consists of yearly modules. Each module is equal to 3 or 4 conventional university-level lessons, depending on the module's difficulty. The students of the HOU study their material, which normally consists of printed textbooks supplemented by audio-visual or electronic material, from distance. Additionally, they have the opportunity to attend a small number of face-to-face counseling meetings. All students have to deliver a number of written assignments, which give them the right to participate in the final examinations held in the end of the academic year. Students attending each module are divided into small groups of approximately 30 students supervised by one tutor. This tutor is responsible for communicating with and supporting each group of students. The module's coordinator bears the overall responsibility of organizing the module, supervising the tutors and defining –in cooperation with the tutors– the educational methods and material as well as ways for supporting the students.

The role of a HOU tutor is quite different from the role of a conventional university professor, as the HOU tutor needs additional skills related to Open and Distance Learning (ODL). An ODL system enables students to control the learning procedure to a great extent, namely choose the place, the time, the procedures or even the material that they will use for studying (Race, 1995). Consequently, the tutor should have an in depth knowledge of the teaching process and of ways to support his students, solve their problems, guide them, or facilitate, manage and evaluate their learning process; this includes the tutor's ability to organize study groups with regard to a specific learning objective (i.e. to organize seminars in a specific programming language such as C++, if the tutor is responsible for the Informatics' module) and to reflect the theoretical approaches on practical experiences (Beijaard and Verloop, 1996). He/she should also help students focus on important subjects and turn their theoretical knowledge into practice.

In addition to the afore mentioned characteristics, the efficiency of a tutor is related to his/her ability to assist students when required, and instruct them or answer their questions in an understandable way. The tutor should also be able to support students even when they are facing personal problems that seem to affect their studies and try to provide solutions. "He should be not too lenient, but remain fair" (Fuhrmann-Greimel and Geyer, 2003). Furthermore, as Villa and Calvete (2001) state, "competence, interpersonal perception, satisfaction, taking risks and initiatives, self acceptance and relationship with students" are six dimensions that must be taken under consideration when defining a tutors' and generally "a teachers' self-concept from a multidimensional perspective".

The role of a module's coordinator should also be further analyzed. In the HOU, all module coordinators belong to the University's permanent staff –as opposed to tutors who sign / renew their contract with the HOU every year. The duties of module coordinators include the organization of the module and the provision of assistance to tutors in understanding the learning process and the particularities of ODL. The coordinator is also responsible for assigning tasks to tutors based on their skills and experience. For instance, a tutor may be an excellent 'conventional' teacher, but if he/she has no experience in ODL and in supporting students form a distance, then he/she should be treated in a different way than an experienced tutor. Therefore, the coordinator's duties include the responsibility of evaluating tutors based on their past activity and renewing their contracts, always keeping in mind that a 'good tutor' is a disputable term and that there exist many crucial factors in a tutor's 'behavior' that –as Ellsworth and Monahan (1998) notice– "are elusive".

In many cases, a tutor's teaching and supporting abilities are 'measured' by his/her own students and many universities consider this type of evaluation as the most influential (Kwan, 1999). Despite the fact that this type of evaluation has become a routine in many universities and students sometimes need to be motivated to have a meaningful and active participation in the evaluation (Chen and Hoshower, 2003), this procedure has been adopted by the HOU since its establishment and forming part of the every-day communication between the students and the HOU. Moreover, once in the academic year a more formal evaluation based on students' opinion of their tutors is held using anonymous questionnaires with 'open-type' questions that students are asked to fill in; the analysis of the questionnaires is usually made using OMR and is subject to the Greek legislation protecting 'personal' data. The results derived from this evaluation are of great importance for the HOU and are processed by the relevant Course. In similar cases, a model is required for automating part of the evaluation process so that it is conducted in an efficient way.

Complementary to this type of tutor evaluation, very often the coordinator also evaluates his/her tutors in many ways. This procedure is usually conducted by hand, i.e. by maintaining a file that helps the coordinator to keep track of the tutors' performance and results in the various tasks and duties that they have been assigned. It is self-evident that the smaller the number of tutors supervised by a coordinator, the less is the time required for their evaluation and instruction. In the HOU case, tutor evaluation is twofold: i.e. tutors are evaluated based on students' data and on the coordinator's data. Since the number of tutors supervised by one coordinator is often quite large, partial computer-assisted automation of the evaluation process is highly desirable. Such automation can be achieved with the use of a model that supports the coordinator by analyzing both students' and coordinator's evaluation data; this is exactly what the presented model does. For example for the course of Informatics / module 'Introduction to Informatics (INF10)', the number of tutors for the academic years 2003-2004, 2004-2005 was 28 and 31 respectively, while in the current year (2005-2006) is 40. In this case the proposed model has proven to be very valuable in automating the evaluation procedure by collecting and graphically presenting data, as well as assisting the coordinator in understanding the tutors' abilities and foreseeing possible problems related to the educational procedure.

An introduction to Bayesian Networks

The model presented in this study is based on Bayesian Networks, also known as Belief Networks, Causal Probabilistic Networks, Causal Nets and more. Bayesian Networks are graphical models introduced by Lauritzen and Spiegelhalter (1988), Pearl (1988) and studied further by scholars such as Jensen (1996, 2001) and Cowell *et al.* (1999). To date, they have been used in the past as tools which can provide valuable help in solving problems in biomedicine and health-care (Acid *et al.*, 2004; Getoor *et al.*, 2004; Lucas *et al.*, 2004), resource and environmental management (Varis, 1998), diagnosis of failures and risk assessment of complex systems in high stakes environments (Neil *et al.*, 2001), financial operational risk scenarios (Neil *et al.*, 2005) and software quality measurement (Fenton *et al.*, 2002). However, *the Bayesian approach in the higher educational context is relatively scarce*, apart from their reported use in the HOU to predict and assess students' behaviour (Xenos, 2003).

According to the graph theory, a Bayesian Network is a directed acyclic graph, in which the nodes represent variables and the directed arrows the connection between them (Jensen, 1996). Therefore a Bayesian Network is a graphic network that describes the relations of probabilities between the variables. The use of this type of modelling brings reason in conditions of uncertainty and combines the advantages of intuition and experience with those of a mathematic model. It provides a powerful framework for modelling of uncertain interaction among nodes (variables) in a specific domain. These interactions can be represented not only in a qualitative manner –via the directed acyclic graph representation– but also in a quantitative manner, by consistently estimating the way in which the initial probabilities influence uncertain conclusions. In this way, the use of Bayesian Network is called *forward estimation*. On the other hand, BN can be used in order to form opinions –without having complete knowledge– about the initial nodes of the network, given the final and some intermediate variables. Their utilization in this case is known as *backward assessment*, for the specific use of BN is for the assessment of an existing state.

The relationships between the variables are defined by a two stage process: firstly, the prior probabilities of all root nodes (nodes with no predecessors) must be given, which are the outcomes of prior research, e.g. in our case data stemming from previous tutors' assessments, or from experience. Secondly, the conditional probabilities of all non-root nodes (nodes with predecessors) given all possible combinations of their direct predecessors are calculated. These are the dependent probabilities that describe the relations between a child node and its parent nodes. In order to illustrate the relations between a 'child' node and its 'parents' in a Bayesian Network, it is necessary to define a way of combining the related probabilities between the variables. Therefore, if a variable has distinct values, a structure called Node Probability Table (NPT) is introduced. This structure is used for representing the probability that a child node is assigned a certain value for each combination of possible values of the parent nodes (Kschischang *et al.*, 2001). For instance, figure 1 represents an example of a Bayesian Network, which consists of two parent nodes (nodes B and C) and one child node (node A). The probability table of node A (NPT_A) reflects the probability P (A| B,C) for all possible combinations of A, B, C. Suppose there are two possible states for node B (b1, b2), two possible states (c1, c2) for node C and three (a1, a2, a3) for node A, then the NPT of node A will include 2*2*3 = 12 elements.

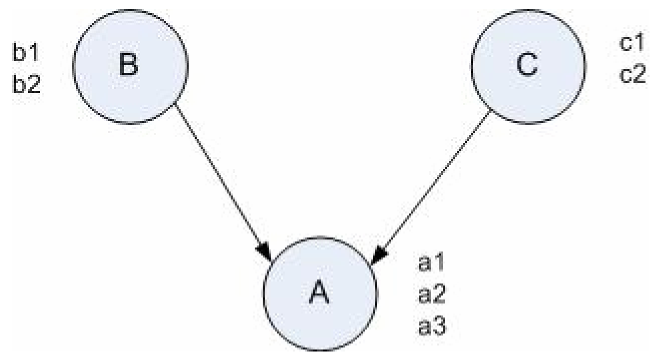

Figure 1. Example of building a Bayesian Network

The mathematical foundations of Bayesian Networks were defined by the mathematician and theologian Thomas Bayes, who based his theorem on the rule that joint probabilities are not as important as conditional probabilities. Equations (1) and (2) are the basic rules of the probability theory. Joining together (1) and (2) results in *Bayes' Theorem*, as shown in (3).

$$P(A \mid B) \cdot P(B) = P(A,B) \qquad (1)$$

$$P(B \mid A) \cdot P(A) = P(A, B) \qquad (2)$$

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)} \qquad (3)$$

If P(A) is a probability about an event A and B provides better information with regard to the event A, P(A) is called *prior probability* and P(A|B) *posterior probability*, representing the probability of an event A, given evidence about how event B affects A. In addition, P(B|A) is the probability that an event B will occur, given a true hypothesis A. The probability P(B) is known as *normalization factor* and does not depend on event B. The advantage of Bayes' Theorem is the fact that probability P(A|B) is easy to be computed in combination with P(B|A). It must be stated that the aforementioned probabilities are hypothetical and determine the confidence level of estimation under certain conditions. Thus, Bayes' Theorem is based on the previous analyses of the probabilities of these conditions. Having in mind Bayes' Theorem, the BN can be used in order to monitor the changes of a case A in relation to a new event B. That is to say that P (A|B) can be calculated by multiplying prior probability P(A) by the probability P(B|A), which represents the event: "B will occur if A is true".

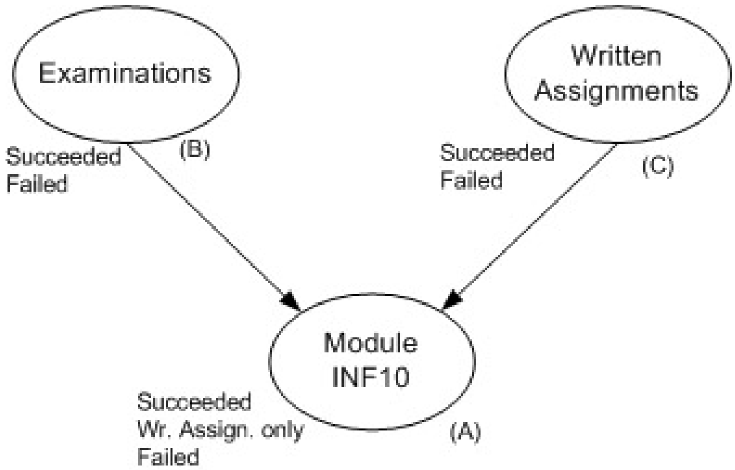

Figure 2. Example of a simple Bayesian Network

Building a Bayesian Network in real-world is quite easy and, according to Fenton *et al.* (2002), can be accomplished in two steps: firstly by constructing the graph representing the random variables as nodes, their correlations as directed arrows and then by assigning probabilities to each of the nodes. Figure 2 presents an (oversimplified) example of a simple Bayesian Network, created to model the behavior of the students of module INF10 of the Hellenic Open University during an academic year. Of course, to define a student's academic behavior numerous factors must be taken into consideration. However, for simplicity reasons only two factors are present in the example's network and are represented in the Bayesian Network by the nodes 'Final Examinations' and 'Written Assignments'. In order to build the specific Bayesian Network, we follow the two-step procedure, as described above. We construct the graph representing the factors that determine the completion of a module in the HOU, that is the nodes 'Final Examinations' and 'Written Assignments' (parent nodes), the final (child) node 'Module INF10', and we draw their correlations as directed arrows. The assignment of probabilities to each of the nodes is determined by the fact that in the HOU, if a student succeeds in both factors (presented by the corresponding nodes in the Bayesian

Network), he will complete the module; if the student succeeds only in 'Written Assignments', he can repeat the examinations in the following year and if he fails again, he must repeat the module; if the student fails in the written assignments, then he/she fails the module and has to repeat it next year. For simplicity purposes, two states named 'succeed' and 'failed' are assigned to both parent nodes and three states named 'succeeded', 'Wr. Assign. Only' and 'failed' to the child node. The graph of this example is shown in Figure 2.
After building the graph of the model, the NPT must be built, taking under consideration the HOU regulations: if the student succeeds in both the final examinations and written assignments, the possibility of success is 100%. If he/she succeeds in the written assignments and fails in final exams, he/she has to sit for exams again and thus the possibility of success is only 50%. If he/she fails in the assignments then he/she is not allowed to sit for exams and thus the possibility of failure is 100%. After assigning all values, the NPT of the sample Bayesian Network will look as Table 1.

Parent nodes		**Child node**		
Final Examinations	**Written Assignments**	**Succeeded (%)**	**Failed (%)**	**Wr. Assign only (%)**
Succeeded	Succeeded	100	0	0
Succeeded	Failed	0	100	0
Failed	Succeeded	0	0	100
Failed	Failed	0	100	0

Table 1. An example of NPT values

In the example's Bayesian Network, when evidence exists that a student has succeeded both in the final examinations and written assignments, the possibility of success in module INF10 is 100% (*forward estimation*). Going backwards, if there is evidence that the student has failed in module INF10, then it is possible that he has failed either in the written assignments, or in the final examinations (*backward assessment*).

To summarize, BN offer the opportunity to define and test hypotheses about an event, due to their advantage in modeling the development of a project where different scenarios exist and state to which extent a scenario will influence the result of the project, as well as to analyse the consequences of the final event to each alternative. Furthermore, BN can be used for taking measures of improvement, if the final events do not meet the initial goals and expectations; they can also be used for analyzing the reasons that led to the undesirable state, so that 'wrong' scenarios will not be followed in solving future problems. The presented study supports that *these c*haracteristics of BN make them suitable for utilization in tutors' evaluation.

Design and implementation of the Bayesian Network

Design of the Bayesian Network

As every educational institute, institutes that are based on ODL aim at offering high quality education to their students. In ODL, as already mentioned, the tutor holds a leading role, acting as a link among the students and the institution or, in other words, as a 'receiver' of all problems that students face during their studies. Thus, it is imperative for every ODL institution to apply a methodology that enables supervisors to identify the skills and / or weaknesses of the tutors, so as to provide help whenever needed and exploit their abilities as much as possible to the benefit of students. Tutors should be made aware of successful educational practices that match students' level and needs, including the specific reasons / circumstances that made their actions successful, in order to be able to repeat and improve them. Similarly, a tutor must be made aware of the factors that lead to unsuccessful educational approaches, so as to be able to avoid taking similar action in the future.

In the HOU case the methodology followed is mainly based on the evaluation of tutors by their students and the coordinator. Comments that could be of use to this evaluation procedure are collected during any form of communication between the students and the HOU. Additionally, an anonymous questionnaire which students are asked to fill in and return to the tutors' supervisor is sent out to students once a year. Then all data are processed, paying attention not to violate the legislation concerning private data protection. Tutors are informed only about the outcome of the evaluation, which is a combination of the students' and the coordinator's evaluation.

The method presented above enables direct evaluation of tutors, i.e. by providing grades that classify them according to 'performance scales'. The presented model acts complementary to this method in an attempt to identify in detail specific factors affecting tutors' performance, i.e. provide more information about their skills and weaknesses than a simple 'Successful' or 'Unsuccessful'. It enables the coordinator of the module or the tutor himself to make assumptions, to experiment on the potential impact of an educational approach and to gain indications of the possible results. Additionally, the model enables the analysis of those factors that influenced a tutor's specific academic 'behavior' in order to answer questions concerning the reasons why a certain decision did not lead to the expected results and define the measures required for improvement. Bayesian Networks are used in this case to assess the probabilities of the NPTs of the parent nodes and the way in which they affect the 'child' nodes of the network.

For instance, the assumption that a tutor is more skilled in giving classroom lectures than in sending supporting e-mails to students can be analyzed by assessing the probabilities of the NPT of the parent nodes referring to communication with the students. Such knowledge can lead the module's coordinator to assign more classroom duties to tutors skilled in face-to-face communication and more communicative and supportive duties to others skilled in writing supporting mails.

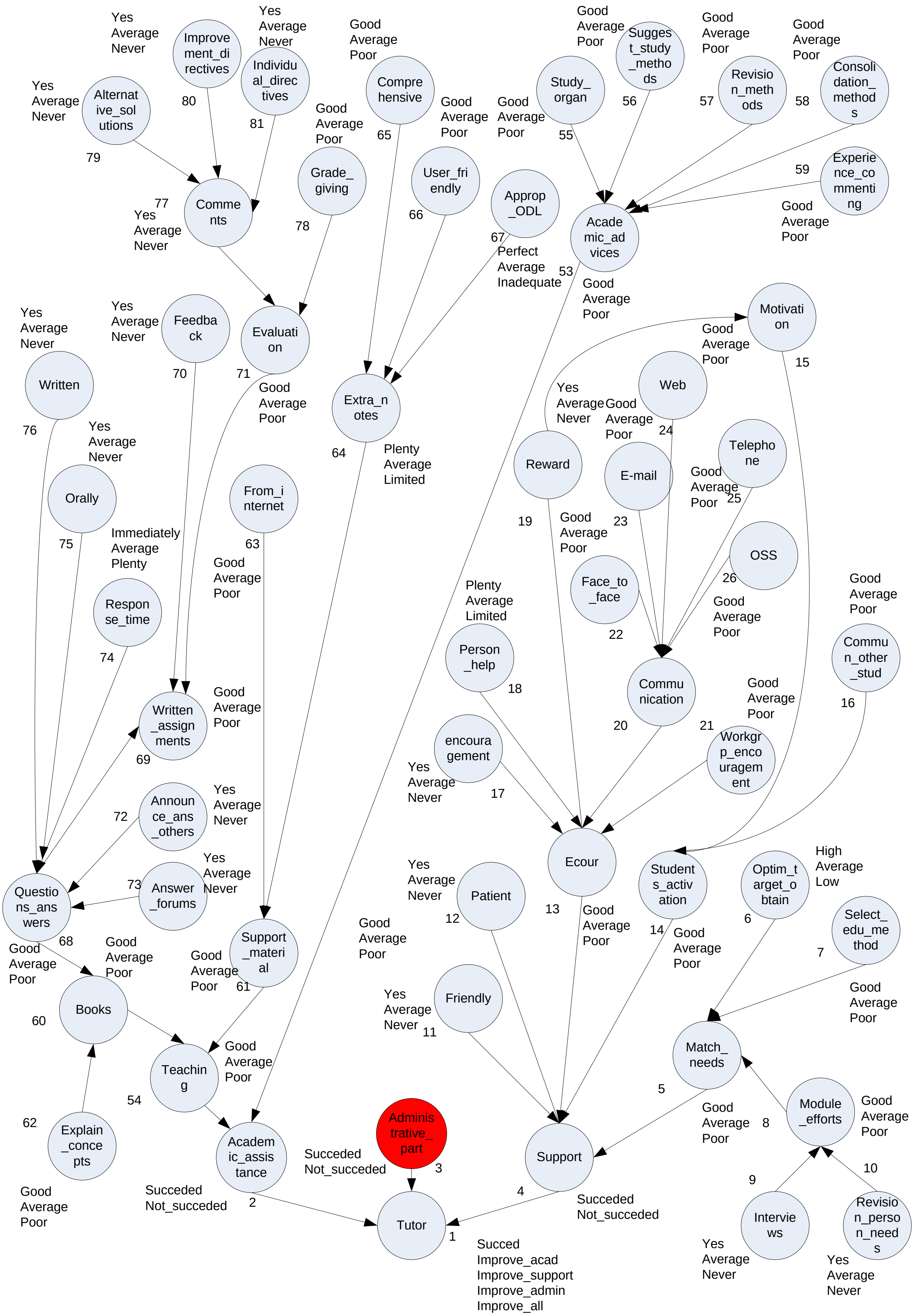

Figure 3. The analysis of node "Tutor" into its parent nodes

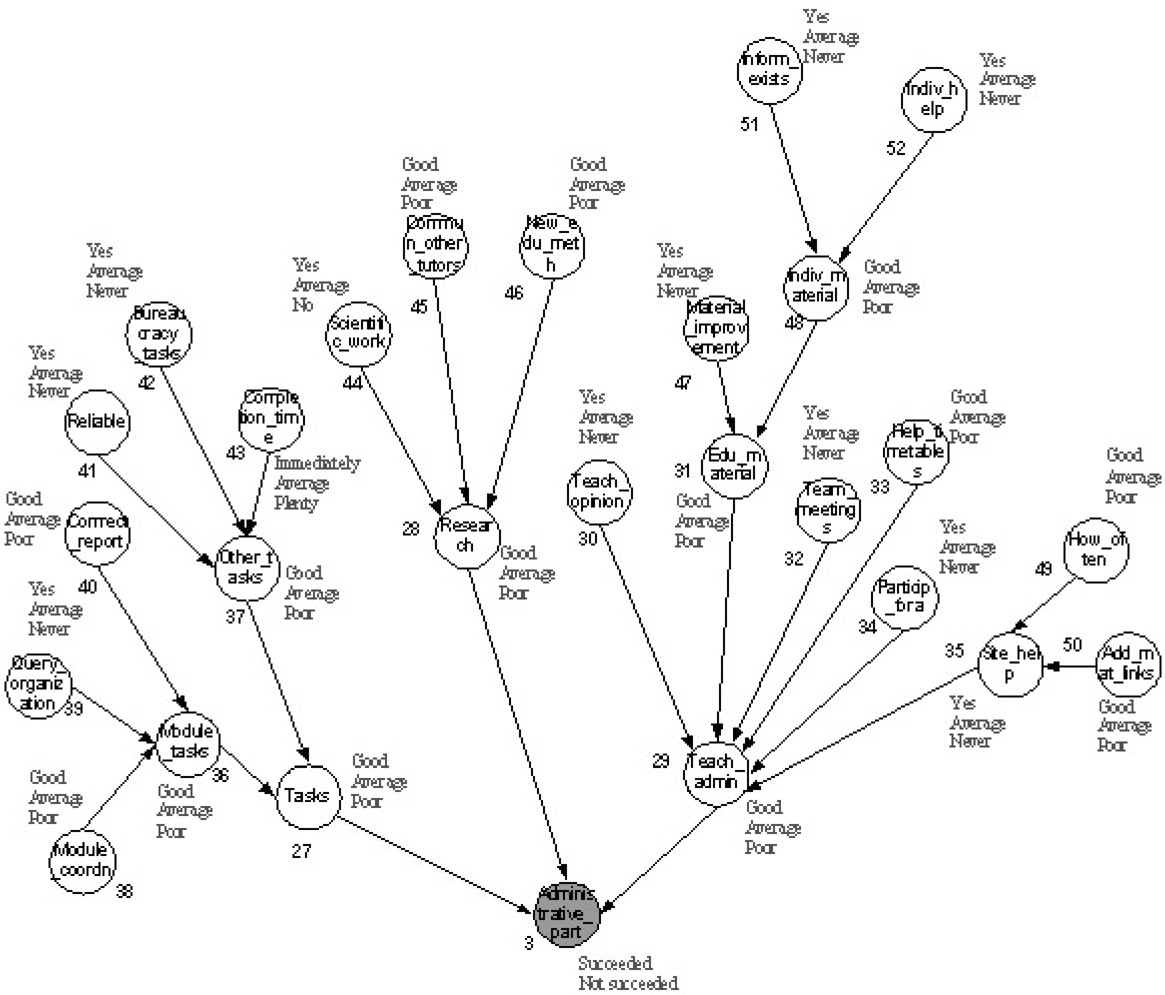

Figure 4. The analysis of node "Administrative Part" into its parent nodes

The presented Bayesian Network is based on the specificities of ODL and the experience obtained from evaluating tutors of the HOU over the last years. The primary goal of the presented effort was to design a network that would take under consideration two important parameters. Firstly, the experience gained during past years available in the form of statistic data collected by various HOU surveys and research, and secondly, data produced by the model itself during the present and future academic years. The model presented in figure 3 is designed to evaluate a tutor's performance during a given academic year. For presentation reasons, the node *Administrative part*, colored in grey, is analyzed in figure 4. The illustration of figure 3 makes it possible to distinguish the final and the intermediate nodes as well as their correlation. Every node of the network represents a variable that is important for tutor evaluation, together with its corresponding states used for building the NPTs of the network. To organize the Bayesian Network in an understandable way, the nodes are arbitrarily enumerated. A 3-degree scale, e.g. *Good*, *Average* and *Poor* is used for most states corresponding to each of the model's nodes, with the exception of specific nodes where it made sense to use the binary scale *Succeeded* and *Not succeeded*. These states do not –of course– imply the definite success or failure of a tutor, but represent probabilities that the specific tutor will be successful or in need of further assistance from the coordinator. The final node of the model has five states. These describe the possibility that the tutor will successfully accomplish his/her academic duties or that further improvement is required in the way that he/she provides assistance and support to students, or in his/her administrative duties, or in all of these factors.

The reason for using a 3-degree state scale in the majority of nodes is that the increase of the number of states leads to increase of the computational complexity of the network, as a larger number of possible values is, in essence, a larger number of probabilities that need to be computed, thus increasing the corresponding computational time. This can be better understood by looking in figure 5 where a simple example of a Bayesian Network is presented; the figure illustrates the increase of the complexity of the NPTs of the node 'Communication' and its 5 'parent' nodes, should the corresponding states increased from 3 to 5 per node. Using a 3-degree scale, the number of probabilities that have to be computed is $3^5 * 3 = 729$. Suppose now that the number of the possible values per node is increased by adding 2 states per node, namely *Very good*, *Good*, *Average*, *Poor* and *Very poor*. In this case, the number of probabilities amounts to $5^5 * 5 = 15625$. Thus, the computational complexity of the NPTs of the network is 'heavily' increased.

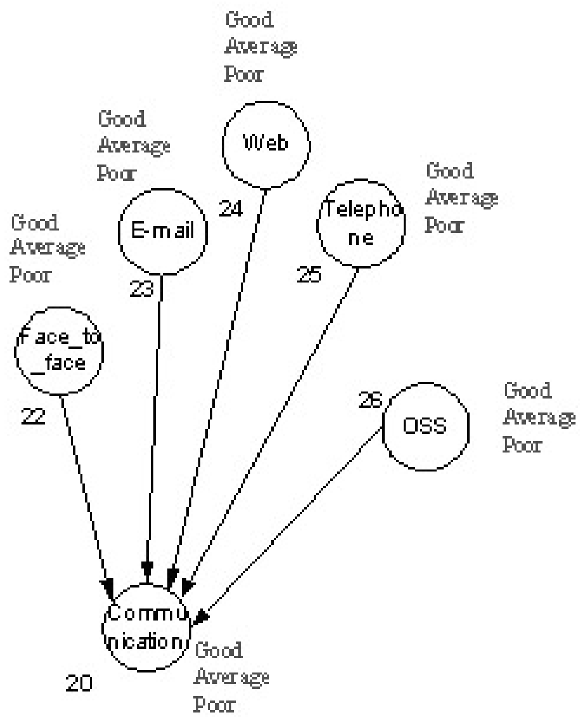

Figure 5. Example of the Increase in the complexity of Node "Communication"

During the design phase of the model, the percentages assigned to each node's states were determined; special care was taken so that these percentages are as realistic as possible and their determination was based on personal experience and statistical data available from previous years. It should be mentioned that these statistical data were the result of research and previous tutors' evaluation by the HOU. The available past data was used to fill in the NPTs of the Bayesian Network, thus enabling the network to provide the initial forward estimations and backward assessments. The process of using past data to calculate the final values of NPT is called 'training' of the model. The initial model was built based on data from the beginning of the academic year 2003-2004. Consequently, as new data became available, the NPT was continuously updated using observations and experiences as well as the results derived from testing hypotheses. At a certain point of this interaction (usually after

4-6 updates, which in our case occurred in the middle of the 2003-2004 academic year) the probabilities reach equilibrium, which means that the deviations in the model's estimations between different steps in the 'training' procedure are minimal and they do not reflect any differences in the estimations. At that point the model is 'trained' and provides accurate results.

As shown in the Bayesian Network of figure 3, a tutor's performance is measured in 3 important fields: *Academic Assistance*, *Administrative Tasks* and *Support*. Of course, there may be other parameters, just as important, affecting the evaluation of a tutor, but in the presented case these 3 are considered as the most important parameters and are thus carefully examined. Specifically, a tutor must offer *Academic assistance* to students. *Academic assistance* refers to provision of assistance to students whenever and in any way needed, as well as provision of supporting material or answering of questions that may arise with respect to the educational material or the written assignments that students have to complete during the academic year. The states that correspond to this node are *Successful* and *Unsuccessful*. The percentages corresponding to each of these states are derived from the values of the NPTs of the parent nodes. Let us take the case where the percentage of 67% corresponds to the node *Academic assistance* (i.e. it is 67% possible that the tutor will succeed in providing assistance to his students). This result is derived from the assessment of the parent nodes' NPTs, namely the provision of assistance, the provision of supporting educational material and the provision of advice to students regarding the organization of their study. To keep the example simple, the nodes presented here are simple and little in number. However, a large number of 'parent' nodes, whose NPTs influence the node *Academic assistance,* can easily be observed in the Bayesian Network.

Another important evaluation parameter, as already mentioned, is the *Administrative Tasks* parameter of a tutor's duties. This node represents the administrative responsibilities of a tutor and is affected by a number of factors –that are represented as nodes in the network– such as the number of tasks completed by the tutor, research on new educational methods, educational material, new and improved ways for communicating with the students, etc.

Finally, the S*tudent's support* parameter is just as important in a tutor evaluation procedure as the two parameters mentioned above and consequently must be taken into serious consideration. In Distance Learning, especially during its initial stages, most students need support for developing the characteristics of a 'self-powered' student. This need for support results from numerous factors, such as the students' lack of 'how-to-study' know-how, as well as their lack of 'self-organization' abilities, which may act as a preventing factor in their efficiently time management. Hence, a tutor should be friendly and patient with students and encourage them in cases of problems that prevent them from studying. Furthermore, among the tutor's duties is the selection of the appropriate educational profile, i.e. one that matches the needs of students. Such profile should function as a motive for students to continue and complete their studies while preventing them from 'dropping out', a phenomenon that occurs in every ODL institution (Xenos *et al.*, 2002).

Following the design of the Bayesian Network and the determination of the specific nodes that it must comprise of, the next step is to implement the network and interpret the corresponding results.

Implementation of the Bayesian Network

The Bayesian Network of the presented case was implemented using *Microsoft © MSBNx Research's Bayesian Network authoring and evaluation tool v1.4.2*. The data inserted to the NPTs of the network were based on previous surveys (Xenos *et al.*, 2002; Pierrakeas *et al.*, 2003), as well as on the experience gained from past tutor evaluations in the HOU. Generally, the implementation of the Bayesian Network is not a complicated task and can be achieved by assessing the NPTs of the network's nodes and then correcting the related probabilities until equilibrium is reached. This task should be done by an expert evaluator, who in the presented case was the coordinator himself. It should be noted that, although it is possible to 'build' the NPTs of the network based on limited information about a tutor's 'behavior', in this case the percentage of uncertainty is quite high and thus the results should be treated accordingly until the model reaches a state of equilibrium. For accurate and reliable results, it is recommended to use data collected from past surveys. Following the completion of the training phase, the coordinator may use the model for feeding in available data and proceed either to the testing of future hypotheses, in which case he/she has no need for knowledge of the BN theory, or to the adjustment of the NPTs.

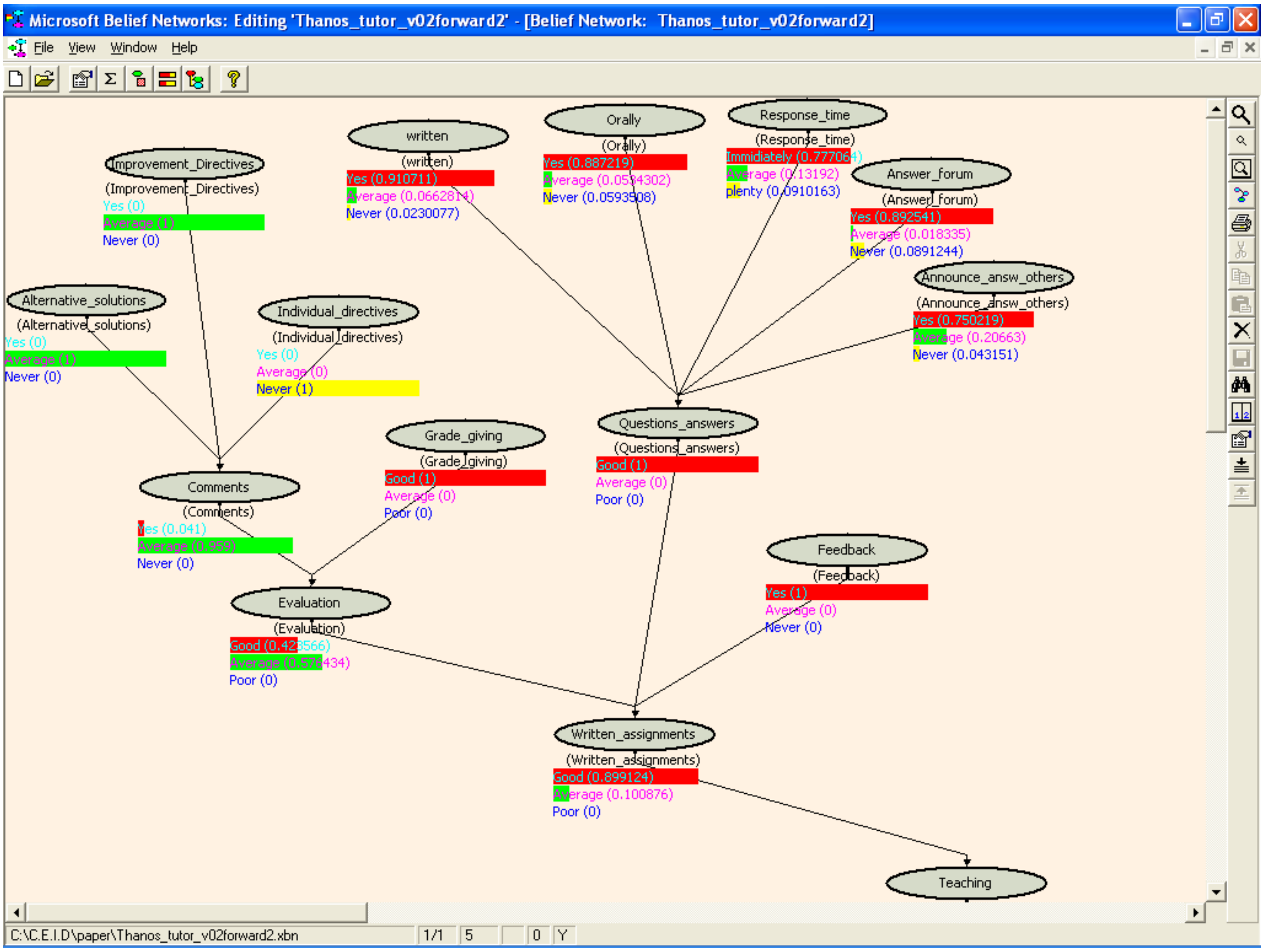

Figure 6. Example of Forward Assessment

The 'prediction' use of the model, in the HOU case, was based on evidence inserted into certain or all of the network's nodes. This data insertion in the NPT of each node resulted in a change of the probabilities and estimations regarding the intermediate and the final node. Every time that new evidence (data) was available, the NPTs were modified accordingly, thus resulting into new and more accurate estimations. This type of implementation is called *forward estimation* and is used to predict possible situations. For example, if the coordinator knows that a tutor is good at providing academic assistance and support during the academic year but is lagging in administrative duties, then he / she can use the model to test possible future reactions of the tutor to a number of solutions and come up with the best solution for each case. For instance, the coordinator might want to simplify some of the tutor's administrative duties or replace them with easier ones that 'match' the tutor's 'philosophy' towards administration. Or the coordinator might decide to provide assistance and advice with regard to those administrative tasks that need to be performed better. Thus, the model can help to test all these cases and predict possible tutor behaviors.

Figure 6 presents an example of use of the Bayesian Network for *forward estimation* concerning node 67 of figure 2 namely *Written Assignments* and its parent nodes. In figure 6 it is possible to observe the nodes and their probabilities. In the HOU case, this is a very important node that greatly affects the completion of a module and the continuation of students' studies, as successful written assignments enable students to participate in the final examinations in the end of the academic year. The node *Written Assignments* corresponds to the tutor's attitude towards students and written assignments and estimates how successful is the tutor or how many written assignments-related tasks need further improvement. It is almost self-evident that a number of factors, other than the simple grading of the assignments needs to be evaluated. In ODL students should be aware of the reasons why they receive the specific grades, which is usually achieved by providing feedback to them. In addition, every corrected written assignment must include comments, alternative solutions and advice for improvement addressed to each student separately. The tutor must also answer a number of questions concerning his / her students' written assignments, the problems faced and their solutions. As illustrated in figure 6, all these tasks constitute the parent nodes of the node *Written Assignments* and their NPTs contribute to the estimation of the probability of the tutor's success.

Prior to the insertion of any evidence the example model estimates a probability of 70% that the tutor will face no problems in the *Written Assignments* task; the remaining 30% is distributed between 28% for average success and 2% that the tutor should significantly improve his / her performance. These estimations are typical percentages derived from past HOU surveys that were used as initial input to the NPTs of the Bayesian Network. Now, suppose that there is evidence that the tutor provided alternative solutions and improvement directives with average success, individual directives with low percentage of success and that there existed no problems in the grading task. This percentage, combined with the successful percentages of feedback and answers to questions gives a final estimation of 90% for *Written Assignments*. It is noted that this percentage was derived automatically, by calculating the probabilities which correspond to the NPTs of all parent nodes of *Written Assignments* node.

Keeping in mind the final estimation, the coordinator or the tutor himself can refer to specific nodes to test possible scenarios. For instance, the percentages of the nodes *Alternative solutions* and *Improvement Directives* were average while the percentage of *Individual Directives* was low (i.e. the directives provided to each of the students were not appropriate). If the percentage of success in all of the above nodes was 100%, the estimation concerning the success probability of the final node *Written Assignments* would be 99%. In another case where the tutor would provide directives for improvement and alternative solutions with average success, with a 100% percentage of success in all other nodes, the success percentage of the final node would be 91%. Therefore, the model enables the coordinator of the module or the tutor to make assumptions about a 'future case' and compute the probability that an event will take place, which aids in reducing uncertainty and occurrence of undesirable conditions.

The presented Bayesian Network can be used not only for forward assessment but also for analyzing the reasons why something did or did not go as planned, which is known as *backward assessment*. An example of backward assessment is presented in figure 6 that refers to node 27 of figure 2, namely to the *Tasks* that a tutor should successfully complete. These tasks are an important part of the Administrative tasks that were regarded, as already explained, as an essential parameter of tutor evaluation during the Bayesian Network design phase. Such tasks include *Module tasks*, i.e. the participation of the tutor in the *Coordination* of the *module*, the *Query organization* and the *Reports* that the tutor has to prepare. In addition to *Module tasks*, a number of *Other tasks* need to be successfully completed such as *Bureaucracy tasks*, the inspiration of *Reliability* to colleagues and adherence to deadlines (*Completion Time*).

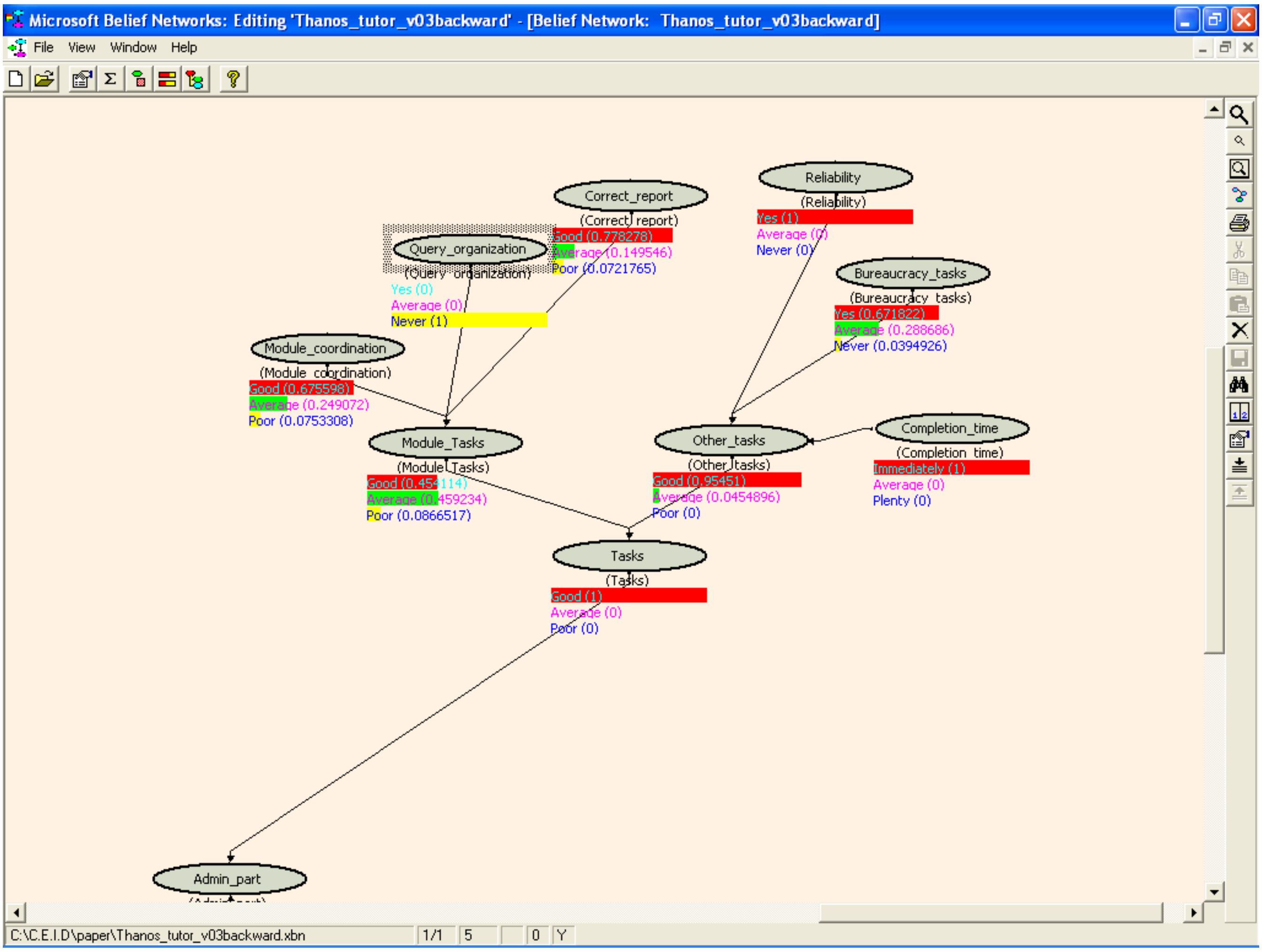

Figure 7. Example of Backward Assessment

In the presented example, an assumption has been made that the probability provided by the model for task completion is 100% 'Average'. This result is derived from assessing the NPTs of the parent nodes of *Tasks*. Suppose that there exist evidence for some of the parent nodes, for instance that the tutor is a reliable person, that he completed his tasks on average time and participated in the coordination of the module with total success. Given this evidence, it is possible to derive results for other parent nodes of the network. As shown in figure 7, the percentage that this tutor will succeed in *Other tasks* is 46%, while in *Module tasks* the percentage drops to 26%; these results provide indication that the specific tutor should improve his performance in both types of tasks. It must be noted that these percentages –where lower than expected– are only *indications* that the tutor might face problems or that he should improve certain aspects of his performance. It is also important to stress the fact that the model's results should always be examined in the context of all tutors' behaviors, namely a tutor's percentage is of no importance, if it is not related to the corresponding percentages of his colleagues.

In order to improve the percentage of the node *Module tasks*, the coordinator or the tutor should carefully examine the parent nodes of the factors that influence the node *Tasks*. As shown in figure 7, the specific tutor should modify his reaction to queries of students or colleagues in order to increase the low percentage. Hence, he should dedicate more time in organizing this duty and consequently reduce the time required for the completion of all other tasks, thus improving the 'child' node *Module tasks*. The tutor may also consider changing the way that the reports are corrected. Specifically, he / she can utilize successful practices of other tutors, or invent new practices and devoting more time in this task. Such action can lead to increase of the success percentage and, as a result, to the better performance of module tasks.

In the case of the *Other tasks* node, the tutor could develop new ways of handling tasks. Successful practices may be proposed to other tutors as well, so that no problems arise again regarding these tasks in the future. In addition, the success percentage of *Bureaucracy tasks* may need to increase, as these tasks affect the final nodes *Tasks* and *Administrative Tasks*, as derived from the requesting help from more experienced tutors, or even from the coordinator of the module.

Summarizing, the use of the Bayesian Network gives the opportunity to predict (*forward estimation*) or to analyze (*backward assessment*) the reasons for the occurrence of a 'desired or non desired' situation. In the presented case, the network provides to the coordinator or the tutor a tool for testing the results of different hypotheses and investigating the reasons that lead to a given state. The Bayesian Network is meant as a tool for partial automation of the evaluation process, optimization of tutors' abilities and improvement of students' studying conditions.

Conclusion

This study presented a method for modeling the evaluation process of tutors in higher education and a model used for the evaluation of the HOU tutors. The introduced model aims not only at making predictions about future 'behaviors' of tutors, but also at analyzing the reasons that lead to the selection and implementation of a given educational action. The presented model has been used in the HOU since 2003.

It must be mentioned that the model does not make any forward estimations or backward assessments 'by itself'. It heavily depends on the data collected from both students and the coordinator, or from surveys usually conducted in institutions of higher education. This data affects the precision of the NPTs of the Bayesian Network and need to be realistic in order for the network to provide accurate results. One of the model's main advantages that differentiate it from other traditional models –which aim only at providing help in marking/grading tutors' performance without any extensive clarification of the factors that led to the specific mark/grade and that had been used from HOU in the past such as questionnaires– is that it can be constantly improved by modifying the NPTs that correspond to each of the factors that determine the tutor's performance. This is done by inserting new data stemming from the final results that the model provides during its utilization. Additionally, the model can easily be expanded by adding new nodes to the network –or altering current nodes– and can be adjusted to the needs of every institution of higher education. The initial effort required to build such a model is significant, however its consequent use is easy and can greatly facilitate the coordinator's (evaluator) effort. Furthermore, since the model is modular, the first version of the network can be easily built and then improved versions can be produced based on new or updated data.

The model presented in this paper is used for tutors' evaluation. Its main advantage consists in helping the evaluator to organize data used for the evaluation, make accurate estimations under conditions of uncertainty and contribute to the continuous improvement of tutors in higher education. A future research goal is the presentation of the results of the evaluation conducted with the use of the proposed model, after collecting and feeding data from at least 5 academic years.

Acknowledgement

The authors will like to thank the tutors, students and administrative personnel of the Hellenic Open University for their assistance during this research effort, as well as the anonymous referees for the valuable comments and suggestions that aided in improving this paper.

References

Acid, S., De Campos, L. M., Fernández-Luna, J. M., Rodríguez, S., Rodríguez, J. M., and Salcedo, J. L. (2004). A comparison of learning algorithms for Bayesian networks: a case study based on data from an emergency medical service. *Artificial Intelligence in Medicine,* 30 (3), pp. 215-232

Beijaard D. and Verloop N., (1996). Assessing teachers' practical knowledge. *Studies in Educational Evaluation* 22 (3), pp. 275-286.

Coffey M. and Gibbs G. (2002). Measuring Teachers' Repertoire of Teaching Methods, *Assessment andEvaluation in Higher Education* 27 (4), pp. 383-390.

Chen Y. and Hoshower L., (2003). Student evaluation of teaching effectiveness: an assessment of student perception and motivation. *Assessment and Evaluation in Higher Education* 28 (1), pp.71-88.

Cowell, R. G., A. P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter (1999). *Probabilistic Networks and Expert Systems. Statistics for Engineering and Information Sciences*. New York: Springer-Verlag.

Ellsworth J' Anne and Monahan A., (1998). Teaching and Learning: Evaluating the novice teacher. *Studies in Educational Evaluation,* 24 (1), pp. 71-86.

Fenton N., Krause P., and Neil, M. (2002). Software Measurement: Uncertainty and Causal Modeling. *IEEE Software July - August.*

Fuhrmann-Greimel B., and Geyer A., (2003). Students' evaluation of teachers and instructional quality – Analysis of relevant factors based on empirical evaluation research. *Assessment and Evaluation in Higher Education,* 28 (3), pp. 229-238.

Getoor, L., Rhee, J.T., Koller, D., and Small, P. (2004). Understanding tuberculosis epidemiology using structured statistical models. *Artificial Intelligence in Medicine* 30 (3), pp. 233-256.

Gibbs G., and Coffey M., (2000). Training university teachers: what are trainers trying to achieve and how could their success be measured? *Teacher Development* 4 (1), pp. 31-44.

Jensen F. V., (1996). *An introduction to Bayesian Networks*. UCL Press.

Jensen, F. V. (2001). *Bayesian Networks and Decision Graphs*. New York: Springer-Verlag.

Kschischang F., Frey B., and Loeliger, H. (2001). Factor graphs and the sum product algorithm. *IEEE transactions on Information Theory, February*.

Kwan K. P. (1999). How fair are student ratings in assessing the teaching performance of university teachers? *Assessment and Evaluation in Higher Education* 24 (2), pp. 181-195.

Lauritzen, S., and Spiegelhalter, D. (1988). Local Computations with Probabilities on Graphical Structures and Their Application to Expert Systems. *Journal of the Royal Statistical Society* 50, pp. 157–224.

Lucas, P. J. F., Van der Gaag, L. C., and Abu-Hanna, A. (2004). Bayesian networks in biomedicine and health-care. *Artificial Intelligence in Medicine* 30 (3), pp. 201-214.

Neil, M., Fenton, N. E., Forey, S., and Harris, R. (2001). Using Bayesian belief networks to predict the reliability of military vehicles. *IEEE Computing and Control Engineering* 12 (1), pp. 11–20.

Neil, M., Fenton, N., and Tailor, M. (2005). Using Bayesian Networks to model expected and unexpected operational losses. *Risk Analysis* 25 (4), pp.963-972.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* San Mateo, Calif.: Morgan Kaufmann.

Pierrakeas C., Xenos M., and Pintelas, P. (2003). Evaluating and improving educational material and tutoring aspects of Distance Learning Systems. *Studies in Educational Evaluation* 29 (4), pp. 335-349.

Race P., (1995). *The Open Learning Handbook. Promoting Quality in Designing and Delivering Flexible Learning*, 2nd edition, Kogan Page, London, 1995.

Scriven M., (1995). A unified theory approach to teacher evaluation. *Studies in Educational Evaluation* 21 (2), pp. 111-129.

Teerajarmorn J., Jamornmann U., and Rowlinson P., (2003). The development of a methodology for scoring university instructors' performance: A case study involving two Thai universities. *Quality in Higher Education* 9 (1), pp. 55-68.

Varis, O. (1998). A belief network approach to optimization and parameter estimation: application to resource and environmental management. *Artificial Intelligence* 101 (1-2), pp. 135-163.

Villa, A., and Calvete, E. (2001). Development of the teacher self-concept evaluation scale and its relation to burnout. *Studies in Educational Evaluation* 27 (3), pp. 239-255.

Xenos M., Pierrakeas C., and Pintelas P., (2002). A Survey on Student Dropout Rates and Dropout Causes Concerning the Students in the Course of Informatics of the Hellenic Open University. *Computers and Education* 39 (4), pp. 361-377.

Xenos M., (2003). Prediction and Assessment of Student Behaviour in Open and Distance Education in Computers using Bayesian Networks. *Computers and Education* 43 (4), pp. 345-359.