Development of an Intelligent Assessment System for Solo Taxonomies Using Fuzzy Logic

John Vrettaros[1,2], G.Vouros[2], and A.S. Drigas[1]

[1] NCSR DEMOKRITOS, Department of Applied Technologies,
Patriarhou Grigoriou, 15310 Ag. Paraskevi , Greece
[2] Aegean University, Info and Communication Systems Eng,
83200, Karlovassi, Samos, Greece
{dr,jvr,imm}demokritos.gr, georgev@aegean.gr

Abstract. In this paper is presented a modeling of assessment systems of taxonomies using fuzzy logic. Specifically the taxonomies system solo is studied, which can be applied in a wide range of fields of diagnostic science. In what concerns education, the test correction is extremely hard and demands experts that are not always available. The intelligent system offers the opportunity to evaluate and classify students' performance according to the structure of the observed learning outcome, concerning the cognitive development of the students in the field of mathematics. The system was tested on high school and university students.

Keywords: fuzzy logic, intelligent system, neural network, assessment system.

1 Introduction: Intelligent Systems and Problems of Knowledge Assessment

An intelligent system is based on an extended quantity of knowledge related to a certain field of problems. This knowledge is organised as a set of rules that allows the system to inference based on the available data. This knowledge –based methodology used in problem solving and more generally in system design has been an evolutionary change in Artificial Intelligence. The consequences of this change are very important since the traditional form of a program (data + algorithm = program) has been replaced by a new architecture, which has as its core a knowledge base and an inference engine under the form:

Knowledge + Inference = System

The specific problem that we have to solve is the construction of an intelligent system, which will be able to evaluate and classify student according to some features, which will be extracted from their answers, into different levels of knowledge. The results are based on a research carried out on high school students and related to the wider field of Mathematics. The classification problem of educated people in different knowledge levels, the study of the transition between these levels

K. Mellouli (Ed.): ECSQARU 2007, LNAI 4724, pp. 901–911, 2007.

as well as the notional change that takes place when a student stops using a naïve (wrong) model and starts using a scientific (right) model, are three of the most important problems in Cognitive Science. A great number of researchers have proposed different methodologies for knowledge acquisition in different scientific fields (Maths, Physics, etc) based on computational and Artificial Intelligence models[8]. Artificial Intelligence methodologies present great interest in theoretic level since they can deal effectively with complexity and fuzziness, which are two of the most important problems in system theory, strongly bound to reality.

In this specific application, analysis starts with the processing of the answers to carefully selected and formed questionnaires which are filled by students. Certain features are extracted out of this analysis that lead to the classification into levels of five different theme sections: Arithmetic, Algebra, Applications Space Perception, and Probabilities and Data. Next, based on this analysis and rule-based knowledge the student classification takes place. Basically the problems that needs to be solved is the automatic classification of students in different levels, using fuzzy logic and artificial neural nets techniques and aiming at creating a system that unifies symbolic and arithmetic processing. For further research, we could note the use experts' knowledge in order to improve the knowledge of educated people (which means transition to a higher level), study the dynamic evolution of the population of educated people and model the changes that take place. Based on the fact that the problem to be solved is a assessment problem, for which there is no specific theory and its data enclose uncertainty (the problem is not purely computational, there is no mathematical solution and the data are not completely known), we can say that use of an intelligent system is appropriate and leads to the construction of a useful tool for student classification in different levels.

The questionnaires that are filled up by students include the aforementioned five theme sections. Each theme section includes four questions, each one corresponding to one of the following levels of knowledge: Single-Structural (SS), Multi-Structural (MS), Relational (R) and Abstractive(S). It should be noted here that the question that corresponds to the Abstractive level cannot be answered by students of the certain age, and consequently we can say that each theme section has three questions. In addition, if a student does not answer any of the three questions in a theme section, he/she is classified in the Pre-Structural (PS) level.

2 Description of the SOM Algorithm Grading System Modelling

The aim of the automatic grading system is the simulation of the teacher's grading system. The answers of the students to the five theme sections are decided into two different categories: Controversial Answers (CA) and Non –Controversial (NCA). Non-Controversial answers are the answers we can be based on in order to classify the student in a level without any uncertainty [1]. For example, if a student gave the following answers to the section that corresponds to Algebra:

Q4. Wrong, Q5. Wrong, Q6. Wrong

Then the student is classified into the Pre-Structural level in Algebra without any controversy. It his/her answers are:

Q4. Right, Q5. Right, Q6. Wrong

Then the student is classifies into the Multi-Structural level in Algebra, without any controversy again. However, there are some answers base on which we cannot conclude to an automatic classification, and we have to take under consideration other factors (in the same way a teacher acts when grading a students answers). For example if a student answers:

Q4. Wrong, Q5. Right, Q6. Wrong

Then his/her classification into a level of knowledge is not straightforward as it was on the examples mention above.

The automated grading system that was developed is illustrated in Fig. 1. The neural network specifies the level of each student in each theme section in cases of Non -Controversial answers. In cases on Controversial answers we have developed two fuzzy systems, since the classification is not obvious and trying to simulate the teacher's way of grading , taking under consideration numerous factors. This way, we take advantage of the symbolic knowledge of system experts and more specific the rule-based knowledge [2]. The first of the two fuzzy systems is implemented based on some statistical analysis and the analysis of some factors such as the Rigour according to which the grading of the certain answer will be done. The second fuzzy system extracts the level of knowledge at each theme section taking into account the Rigour (which the previous system's output) and the answers given to the questions of the specific theme section. Next, the final level is determined for each student based on the results (outputs) of the above systems.

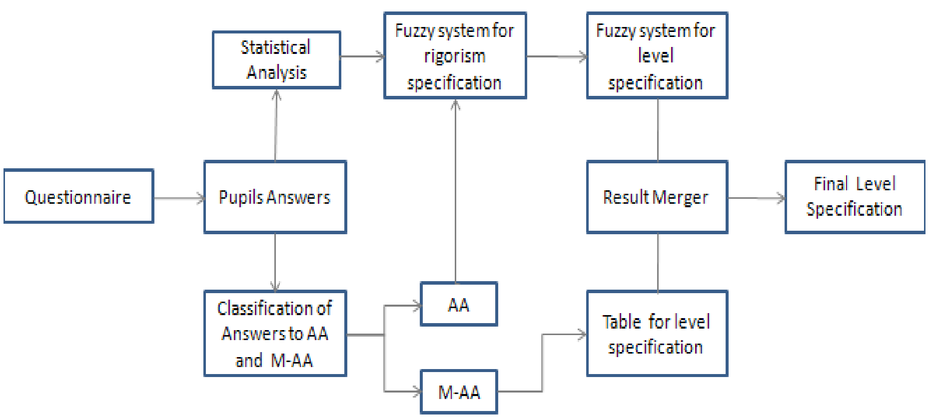

Fig. 1. Grading Modelling System

3 Student Level Determination System for Each Theme Section: Non-Controversial Answers

In the previous section we briefly described the procedure that was followed in order to implement the automated classification of students into levels of knowledge. We mentioned that the students' answers, in the five different theme sections, are divided into two categories: Controversial and Non-Controversial [3]. For the first category, Table 1. was used in order to extract the results. In the specific application, an algorithmic method could also be used for the extraction of the final result. However, based on the fact that we are interested in the extension and application of the developed system in more complicated problems (i.e. we will ask the system to grade the answer using a grade between 1 and 10 or 100), the use of the table is the most appropriate. The Non-Controversial answers are illustrated in Table 1. We have used 0 to symbolize the wrong answer, 1 for the right answer and 2 for invalid answer (case where the student does not answer).

Table 1. Classification based on Non - Controversial answers

SYMBOL	ANSWERS	KNOWLEDGE LEVEL
000	WRONG – WRONG – WRONG	Pre-Structural
100	RIGHT – WRONG – WRONG	Single-Structural
110	RIGHT –RIGHT - WRONG	Multi-Structural
111	RIGHT – RIGHT - RIGHT	Relational
222	INVALID – INVALID - INVALID	Pre-Structural
220	INVALID – INVALID - WRONG	Pre-Structural
202	INVALID – WRONG - INVALID	Pre-Structural
200	INVALID – WRONG - WRONG	Pre-Structural
122	RIGHT – INVALID - INVALID	Single-Structural
120	RIGHT – INVALID - WRONG	Single-Structural
112	RIGHT – RIGHT - INVALID	Multi-Structural
102	RIGHT – WRONG - INVALID	Single-Structural
022	WRONG – INVALID - INVALID	Pre-Structural
002	WRONG – WRONG - INVALID	Pre-Structural
020	WRONG – INVALID - WRONG	Pre-Structural

4 Student Level Determination System for Each Theme Section: Controversial Answers

In the previous section we referred to the cases where the classification of the students into knowledge levels is done based on their answer without any uncertainty. In this section we will refer to the Controversial cases where the student classification in some level cannot be done without any uncertainty [5]. For the evaluation of these answers we will consider the following factors (that correspond to the factor that the teachers take into account when dealing with controversial cases):

1. The difficulty of the certain subject that obviously affects its grading.
2. The number of void answers, which is the number of question that were left unanswered by the student. This factor is considered since it affects the student's evaluation. If, for example, we want to grade a controversial answer (e,g case Q4. WRONG, Q5. RIGHT, Q6. RIGHT) and the student has a great number of unanswered questions, this means that probably the student is not answering the questions randomly, but he/she answers the question seriously. We conclude that probably the wrong answer in Q.4 is wrong due to carelessness, since the right answers in Q.5 and Q.6 (which are obviously much harder to answer that Q.4) are not given by chance. Consequently the student can be classified in the Relational level in the specific theme section..
3. Child level, meaning the general impression the student makes.

Controversial cases occur when a right answer follows a wrong one. These cases are 12 in total, as it is presented in the following table (Table 2):

Table 2. Controversial answers description

SYMBOL	ANSWERS	KNOWLEDGE LEVEL
001	WRONG – WRONG - RIGHT	?
010	WRONG – RIGHT - WRONG	?
011	WRONG – RIGHT - RIGHT	?
101	RIGHT – WRONG - RIGHT	?
221	INVALID – INVALID - RIGHT	?
212	INVALID – RIGHT - INVALID	?
211	INVALID – RIGHT - RIGHT	?
210	INVALID – RIGHT - WRONG	?
121	RIGHT – INVALID - RIGHT	?
021	WRONG – INVALID - RIGHT	?
012	WRONG – RIGHT - INVALID	?
201	INVALID – WRONG - RIGHT	?

In general, we can say that the selection of the level in cases of the controversial answers is different depending on the student. It is affected by the student's answers to previous questions, the number of questions that are left unanswered and the level of the question. In order to model the controversial cases there have been designed and implemented two fuzzy systems, that are analytically described in the next sections [6].

4.1 Rigor Grading Determination Subsystem

The first system evaluates the Rigor according to which the student will be graded. Rigor is a number between 0 and 1 and it is used for the classification of student in knowledge levels. The system has three inputs and one output. The inputs are factors that affect the grading of each controversial answer: number of unanswered question, question level and child level [8]. The output is one: the Rigor. The question level is evaluated according to the answers of other students to this question. The more the students that answered this question, the easier the question is and the Rigor level is increased. The child level is estimated based on the student's answers.

In this specific case the values of each input are between two values. The difficulty of the subject and the number of void answers take values between 0 and 100, and the Child level take values between 0 and 3. The Difficulty of the subject is estimated based on other students' answers. The x axis –s normalized and has values in the range [0,100]. We can define a partition on the domain field of Subject Difficulty, for example we can say that if value of Difficulty is in [0, 30) then Difficulty is small, if it is in [30, 65), then is said to be medium and finally if it belongs in [65,100), is said to be larger. However this way of classical partitioning introduces great uncertainty in some areas (e.g. close to 30, 65 and 100). This means that if the value of Difficulty is equal to 29 then difficulty is small, whereas if it is equal to it is medium. In order to avoid such problems we define the fuzzy partitions (one for each input) on the domain field of each input A_1, A_2 and A_3. Each fuzzy partition is of order 3. A Fuzzy partition B, of order 3, is defined on the domain field of the output, which is [0, 1] as mentioned above. The fuzzy partitions A_1, A_2 A_3 and B linguistic representations of the domain fields and consequently their elements are linguistic terms of the form SMALL, LARGE, MEDIUM, etc. In Fig. 2. the fuzzy partition of the first input is described.

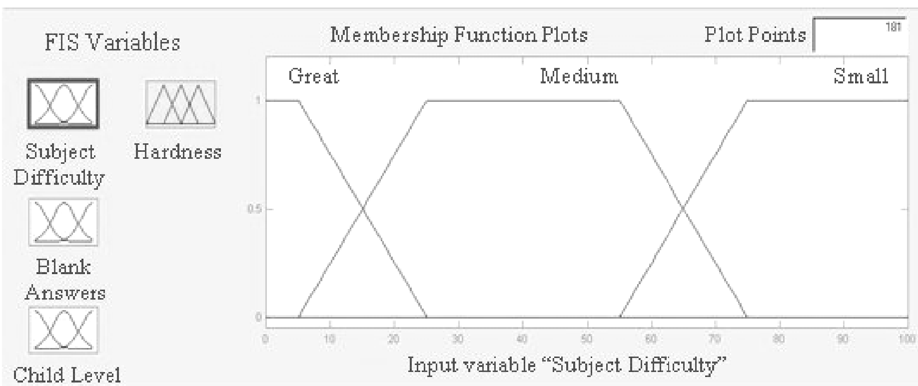

Fig. 2. Fuzzy partition of 1st input: "Subject Difficulty"

If the value of the input is for example equal to 2 or 90, then we can say that the Subject Difficulty is 100% SMALL or 100% LARGE respectively. But if the input value equals 15 then the Subject Difficulty is 0.5 SMALL and 0.5 MEDIUM, and if it is equal to 60 is 0.75 MEDIUM and 0.25 LARGE. The above consideration, that is the use of fuzzy partitions, is obviously much closer to reality since it simulates better the human way of thinking.

The next step is to define the rules of the system. These rules are the following:

1. If Subject Difficulty is "BIG" then Rigor is "HIGH".
2. If Subject Difficulty is "MEDIUM" then Rigor is "ENOUGH".
3. If Subject Difficulty is "SMALL" then Rigor is "LOW"

4. If Void Answers are “FEW” then Rigor is “SMALL”.
5. If Void Answers are “ENOUGH” then Rigor is “ENOUGH”.
6. If Void Answers are “MANY” ” then Rigor is “HIGH”.
7. If Child Level is “LOW” then Rigor is “LOW”.
8. If Child Level is “AVERAGE” ” then Rigor is “ENOUGH”.
9. If Child Level is “HIGH” then Rigor is “HIGH”.

The system’s domain field is shown in Fig. 3.

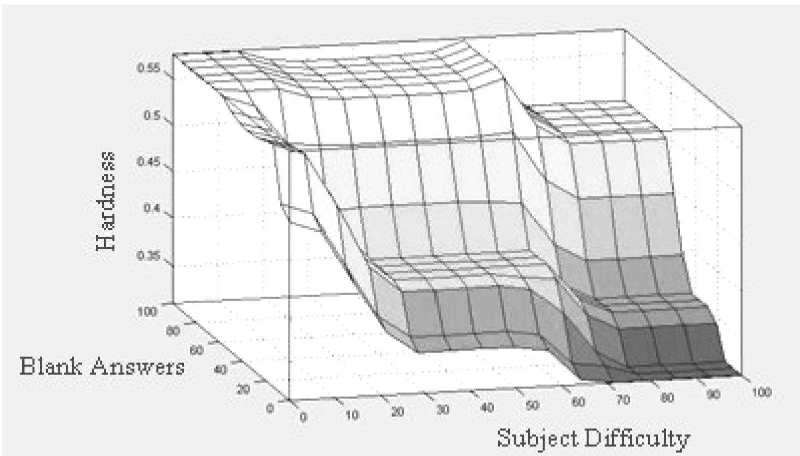

Fig. 3. System’s Domain Field

4.2 Student Level Determination Subsystem Per Theme Section

The second fuzzy system determines the student’s level in each theme section, and more precisely based on the Rigor of grading (that is the output of the previous system) a fuzzy system has been designed student classification in theme sections. In the second fuzzy system, we use only the data of the student under consideration. The two fuzzy systems together with the neural network provide us with the answers of the first part, meaning the student classification in one of the aforementioned knowledge levels per theme section [10].

The system has two inputs and one output. The first input is the Rigor and the second one is the three answers for the subject. The output is a number between o and 3 that corresponds to the four levels of knowledge (Pre-Structural, Single-Structural, Multi-Structural and Relational) for each theme section. The result in controversial cases can be a decimal number.

The values that Rigor is allowed to take are 0.2 toι 0.65. The values of X-axis are covered by two membership functions. If, for example, we take as input the value 0.3, the Rigor is 50% Low and 50% enough. The second input is the answers of the student under investigation. The input’s values are normalized from 1 to 10. In order to interpret the student answers in values between 0 and 10 we apply the following formula:

$$\text{Results} = \alpha + \beta + \gamma + \delta$$

where α is the number of right answers, β is the number of last right answer, γ is the number of first right answer δ is the number of void answers. The controversial answers can only have two values otherwise they are not controversial. ($a \leq 2$). The greatest number we can have is 10 and the least 4.

The second input of the system is the student level in the theme section under consideration. The output of the two fuzzy systems, in combination with the output of the neural network, determines the levels of the students in the five theme sections.

The output values are between 0 and 3.) corresponds to Pre-Structural level, 1 to Single-Structural, 2 to Multi-Structural and 3 to Relational. The rules that associate the inputs with the outputs are the following:

1. If Rigor is HIGH then the Level is LOW.
2. If Rigor is AVERAGE then the Level is MEDIUM.
3. If Rigor is LOW then the Level is HIGH.
4. If the Answers are FEW then the Level is LOW.
5. If the Answers are ENOUGH then the Level is MEDIUM.
6. If the Answers are MANY then the Level is HIGH.

5 Final Level Determination

Up to now, we have determined the levels of knowledge of students in five different theme sections. Based on these levels we will determine the final overall level. The final level is a number between 0 and 3 that corresponds to one of the four knowledge levels. In the previous sections we described the procedure of level determination based on the theme sections. The procedure that follows next investigates the students' answers according to knowledge levels rather than theme sections [11]. The degrees of trust will specify at what point the student under investigation belongs to each level. The degree of trust is number between 0 and 1.

The system was divided into four parts, each one associating the number of given answers to the number that we believe it belongs to the specific level (Fig. 4.).

The degrees of trust are three: one for the Single-Structural, one for the Multi-Structural and one for the Relational. For the Pre-Structural there is no degree of trust because it always equals 1 since there are not any questions or answers and additionally it is the lowest level and consequently there can be no degree of trust less than 1. Next, having 3 degrees of trust we decide on the final level by taking the average. The average is taken according to the level. Having o 1 for the Pre-Structural, 2 for the Multi-Structural and 3 for the Relational we get:

$$\varepsilon = \frac{1C_1 + 2C_2 + 3C_3}{C_1 + C_2 + C_3},$$

where ε is the final level. ε can be a decimal number. For example if $\varepsilon = 1.5$, then the student is uniformly classified between Single-Structural and Multi-Structural Level.

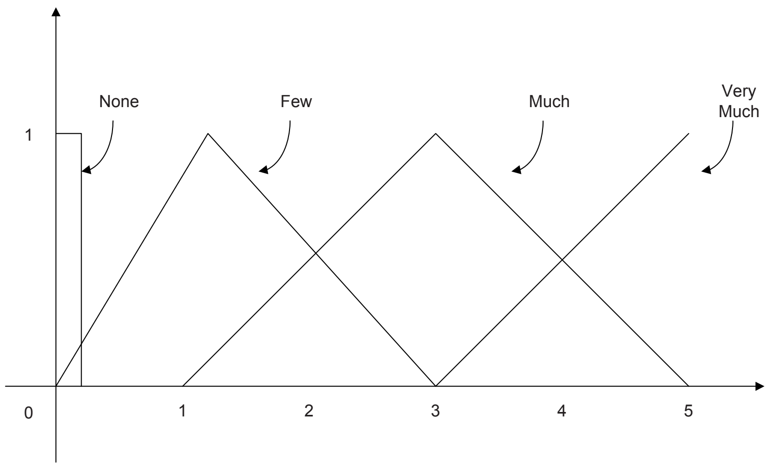

Fig. 4. Number of answers in the specific level

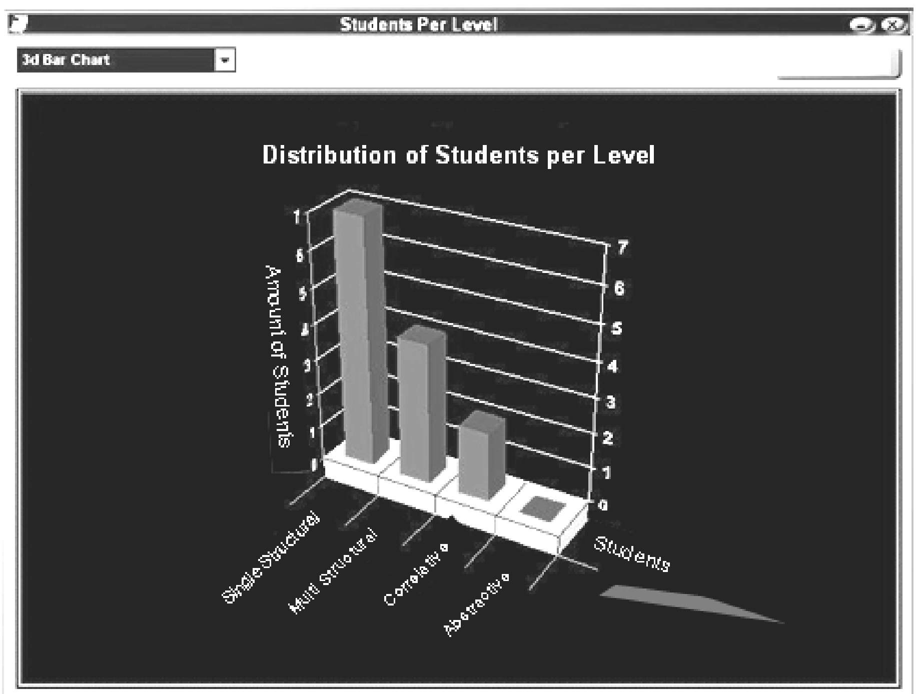

Fig. 5. Students per Level Graph

6 Case Study: The Solo Program

The SOLO program is the interface that contains a powerful intelligent engine that uses an educational diagnostic tool, which basically manages the data of the class and the students [7]. It is very simple and easy to use, providing help support. Below are stated some selections provided by the interface:

Students per Level: With this selection the user is provided in 2D or 3D graph the distribution on the students depending on the level the students are (Fig. 5).

New Database: This function provides to the user the possibility to create a new database. The window contains combo boxes and textboxes where the user inputs the variables. Using the add button the user inputs a new record to the database. With the delete button the user can delete the present record. By pressing the refresh button the user can refresh the database (for multi-user environment only). With the update button the user can post the database for the changes done, and with the exit button the user closes the window and returns to the main window of the application. By pressing the SaveDB the user saves the database to the hard-disk. (Fig. 6)

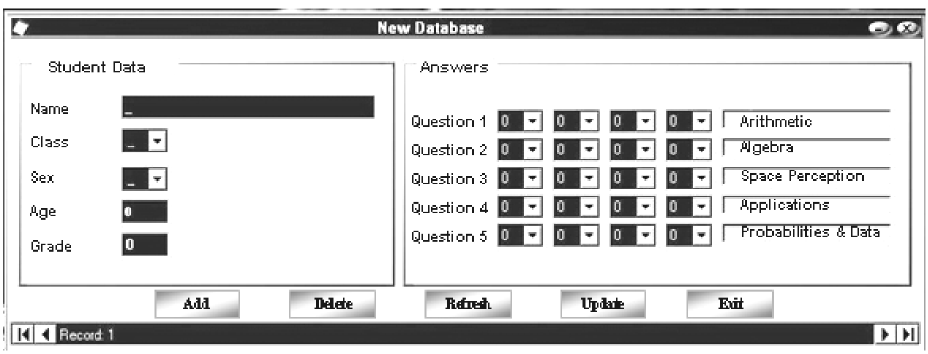

Fig. 6. New Database Window

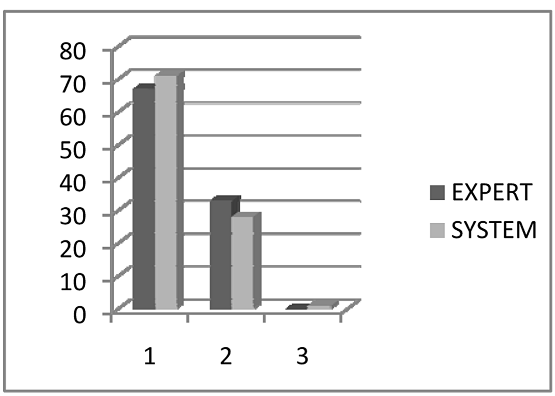

Fig. 7. System -Expert Results Comparison per Level (Dark Grey color corresponds to the expert and Light Gray color corresponds to system)

7 Conclusion

In order to prove the effectiveness of assessment tools, the developed system was applied on 100 high school and senior high school students, and it was tested on mathematics. The correction results obtained by the system were compared to the results obtained by the cognitive science expert. The system's results were found to be very close to the expert results, as it can be seen on the following table (Fig. 7). Concluding, we can say that the assessment tools are trustworthy tools for the educators' cooperation and contribution.

References

1. Biggs, J.B., Collis, K.F.: Evaluating the Quality of Learning: the SOLO Taxonomy. Academic Press, New York (1982)
2. Biggs, J.B., Collis, K.F.: Multimodal learning and the quality of intelligent be- havior. In: Rowe, H.H. (ed.) Intelligence: Reconceptualization and Measurement. Hillsdale, N. J.: Lawrence Erb.freemanm Associates and Hawthorn, Vic.: ACER, pp. 57–76 (1991)
3. Drigas, A., Kouremenos, S., Vrettos, S., Vrettaros, J., Kouremenos, D.: An ex- pert system for job matching of the unemployed. Pergamon-Elsevier Science LTD, Oxford, IDS (2004)
4. Cox, E.: The Fuzzy Systems Handbook: A Practitioner's Guide to Building, Using, & Maintaining Fuzzy Systems (1999)
5. Gogus, O., Boucher, T.O.: A Consistency Test for Rational Weights in Multi- Criterion Decision Analysis with Fuzzy Pairwise Comparisons. Fuzzy Sets and Systems 86, 129–138 (1997)
6. Imrie, B.W.: Assessment for Learning: Quality and Taxonomies. Assessment and Evaluation in Higher Education 20(2), 175–189 (1995)
7. Koleza, E., Barkatsas, A.: The SOLO Taxonomy as a tool for evaluation in Mathe- matics, 2nd Mediterranean Conference on Mathematics Education, Nicosia-Cyprus (2000)
8. Maeda, S., Murakami, S.: The Use of a Fuzzy Decision-Making Method in a Large- Scale Computer System Choice Problem. Fuzzy Sets and Systems 54, 235–249 (1993)
9. Nastoulis, C., Leros, A., Bardis, N.: Banknote.: Recognition Based on Probabilistic Neural Network Models. WSEAS Transactions on Systems 5(10), 2363–2367 (2006)
10. Vrettaros, J., Grigoriadou, M.: Design of hybrid architecture for fuzzy connectionist expert system and its application to approximate student model. In: CATE 96. The first international conference on computers and advanced technologies in education, Cairo (1996)
11. Zadeh, L.F.: Toward a Theory of Fuzzy Information Granulation and its Centrality in Human Reasoning and Fuzzy Logic. Fuzzy Sets and Systems 90, 111–127 (1997)