Gazi University Journal of Science
GU J Sci
24(2):317-327 (2011)

GUJS
www.gujs.org

An Expert System Approach For High School Type Selection

Süleyman ERSÖZ[1], Adnan AKTEPE[1♠]

[1] *Kırıkkale University, Faculty of Engineering, Industrial Engineering Department, 71450, Yahşihan, Kırıkkale, TURKEY*

Received: 13.10.2010 Revised: 23.11.2010 Accepted: 14.12.2010

ABSTRACT

An expert system is a computer program that imitates the procedures by which experts solve problems. This paper describes an expert system application which provides advice to primary school students who are seeking assistance in determining high school type. The system gathers information about student grade point average (GPA), centralized exam score, personality type and socio-economic factors. It assesses student qualifications for seven different school types in Turkey. The expert system recommends a school type for the students and produces a short report explaining the reasons of recommendation. Student GPA and centralized exam score are considered as technical criteria. In addition, social and socio-economic factors such as personality type, parental involvement etc. affecting the decision process are also considered and included in the model.

Key Words: *High School Type Selection, Expert Systems*..

1. INTRODUCTION

Education has been an important factor for economic development and social mobility both industrialized and developing world and it will continue to be a crucial factor in the future [1]. Education is an investment tool that one can benefit from during his/her life. For an equipped education, people need to attend qualified schools. Therefore, school type selection is a crucial factor for this investment. In the literature several criteria have been defined fort this problem. There are basically three strands of literature that address school selection: i) social psychological studies, i.e., parental influence; ii) status attainment, i.e., social status; and iii) economic studies such as cost benefits and rational selection [2].

School selection is a multi-criteria decision making problem. In a decision making problem there are several uncontrollable parameters. Decision modeling studies the human decision making mechanism and tries to build models that predict human decisions [3]. Expert systems can be considered as decision modeling applications. Expert systems incorporate decision-making processes, and can be considered as the mechanization of human thinking [4].

In this study an expert system model is created for high school type selection problem for primary school graduates in Turkey. School selection criteria are grouped into 3 main categories which are technical criteria, personality type criterion and socio-economic criteria. Technical criteria are student GPA and centralized exam score. Personality type is included in the model by using personality score rules after determining certain type of

♠Corresponding author, e-mail: aaktepe@gmail.com

personality with a personality type test. Socio-economic criteria are parental involvement, economic condition of the family and social affinity of the student. In order to consider socio-economic factors a survey is conducted to high school students. The factors correlated with high school type selection are subsumed in the expert system model converting the results into rules.

In the next section factors affecting school selection problem are discussed with related literature. Third section of the study is about the developed expert system application. The paper ends with a conclusion section discussing the properties, use and advantages of the model.

2. HIGH SCHOOL TYPE SELECTION PROBLEM AND THE SITUATION IN TURKEY

To choose a certain type of school firstly academic achievement criteria must be met. Graduating GPA and national placement test score are two main criteria for entering certain type of school around the world. The students who cannot afford high grades generally attend general high schools and vocational high schools. The students who can afford high grades can attend better schools such as specific colleges, science high schools etc. However, not only academic achievement criteria but also other criteria such as personality type and socio-economic condition of the family must be considered in order to determine the best selection.

Personality type is also an important factor in high school type selection decision process. Archbald [5] and Eksi et al. [6] discuss the role of personality in school selection. The personality type of the student must also be taken into consideration in the model. For determining the personality type of the students a personality test is conducted to students in this study. There are several kinds of personality type determination tests. One of them is the Holland Quiz [7]. There are several kinds of the test for different age groups which are available from the website [8]. The main purpose of the test is to determine the personality type. There are six different personality types which are realistic, investigative, artistic, social, enterprising and conventional. In this study, a similar personality type test is conducted to the students. The test is maintained from a psychological counseling center located in Ankara, Turkey. The result of the test in this study gives 6 different personality types which are: realist, explorer, entrepreneur, artistic, social and traditional personality types.

Besides academic achievement and personality type, Aypay [2] lists the factors that lead to students' selection as gender, parent–child discussion of school, parent–school relationship, family income, mother's occupation, father's occupation, guidance in primary school, mother's education, father's education and luck. Parent involvement is an important factor because students make high school type selection decisions just after primary school at ages about 13-14. This considered in several papers in the literature. Coleman [9], Epstein [10], Lee et al. [11], Walker and Clark [12] and Lai et al. [13] discuss the parent involvement in students' school life and their role in their children's decision making about education life. These papers emphasize the possible negative and positive effects of parental involvement. The parental involvement effect on school selection problem is modeled as socioeconomic factor in this study which is discussed in detail in the next section.

In Turkey, there are 7 major types in high school selection which are Science High School, Social Sciences High School, Anatolian High School, Vocational High School, Fine Arts High School, Military High School and Others (e.g. General High Schools). Science High Schools are the best selection for creative and hardworking students and the student having high graduating GPA and high national placement test score can attend these schools. The creative and hardworking students whose placement test score is lower comparing to Science High School base points choose Anatolian High Schools. The students who want to become an officer choose Military High Schools. The students who are interested in art choose Fine Arts High Schools. The students whose diploma degree and placement test score is low and want to gain expertise in several types such as automotive, furniture, computer technologies etc. choose Vocational High Schools. Other students with low graduating diploma degree and low placement test score choose other high schools such as General High Schools. High school education is not compulsory in Turkey now. Therefore, a few of the students may prefer not to attend high school.

Psychological counseling and guidance teachers in Turkey help the students in this decision problem. However, their efforts are not enough, because they cannot allocate enough time for each student and mostly they do not know or measure the personal capabilities of the students. Turkey has a relatively large young population. 13,000,000 students are enrolled in primary and secondary public schools with half a million 500,000 teachers in the Turkish Educational System. Only a few of 500,000 teachers are psychological counseling and guidance teachers and most of the personnel cadre are empty for these teachers in schools. So the psychological counseling and guidance doesn't work well. In order to support students and their families for high school type selection a comprehensive expert system model including both technical and social criteria is developed in this paper.

3. AN EXPERT SYSTEM APPLICATION ON HIGH SCHOOL TYPE SELECTION FOR TURKISH PRIMARY EDUCATION SYSTEM

Expert systems are the computer systems which can model the decision process of an expert or a number of experts [14]. Expert systems have been used to solve a wide range of problems in domains such as medicine, mathematics, engineering, geology, computer science, business, law, defense and education. Within each domain, they have been used to solve problems of different types. Types of problem involve diagnosis (e.g., of a system fault, disease or student error); design (of a computer systems, hotel etc); and interpretation (of, for example, geological data) [15]. Other application areas of expert systems are discussed in Liao's [16] and Bo and Monaco's [17, 18] studies in detail.

This paper presents the application and use of expert systems in an education system. The input parameters of expert system model developed for high school type

selection are graduating GPA of the student, centralized exam score, personality type of the student and socio-economic factors. Next 2 subsections discuss how to incorporate personality characteristics and socio-economic factors into the model.

3.1. Incorporating Personality Type Factor Affecting School Selection Into The Expert System Application

To incorporate personality type effect into the model, a personality type determination system is used. Personality test score is obtained after taking the test. There are 90 questions in the personality test. The test determines the overwhelming personality type. Different

questions measure different personality type. The set of questions each set measuring different personality type are shown in Table 1. The users do not know which question measures which personality. There are 3 selections for each question in the test which are "like" (H+), "do not like" (H-) and "doesn't matter" (F). The user gets 1 point for the selection selected. For example, if he/she chooses H+ for Question.1 he/she gets 1 point for Explorer H+ (ExH+).

Table 1. Personality types and set of questions.

Personality type	**Set of questions**
Explorer	1, 3, 4, 8, 11, 17, 22, 34, 39, 41, 47, 48, 68, 71, 80
Artistic	6, 10, 14, 31, 32, 42, 44, 45, 50, 62, 72, 74, 77, 78, 81
Social	2, 21, 29, 37, 49, 55, 58, 60, 64, 65, 66, 73, 87, 88, 90
Entrepreneur	7, 15, 16, 18, 20, 28, 33, 35, 51, 56, 63, 67, 75, 83, 85
Traditional	5, 12, 23, 24, 27, 30, 36, 43, 46, 57, 69, 76, 86, 89
Realist	9, 13, 19, 25, 26, 38, 40, 52, 53, 54, 59, 61, 70, 82, 84

In order to determine the overwhelming personality type, following are created for calculating certain type of personality test score:

IF H+ points > H- points AND H+ > F points

THEN multiply related personality type score with 1

IF F points < H- points AND H+ points < F points

THEN multiply personality type score with -1

IF F points < H+ points AND H- points > H+ points

THEN multiply related personality type score with –0.8

IF F points > H- and H- points AND H- points >= H+ points

THEN multiply related personality type score with -0.6

IF F points >= H+ and H- points AND H+ points > H- points

THEN multiply related personality type score with 0.6.

After calculating the scores for each type of personality, the final score is calculated. Figure 1 below shows the decision tree of the final personality type determination process with overall scores. Overall H+ score is calculated with rules given above. Overall H+ score for a personality type greater than any other personality type determines the personality type.

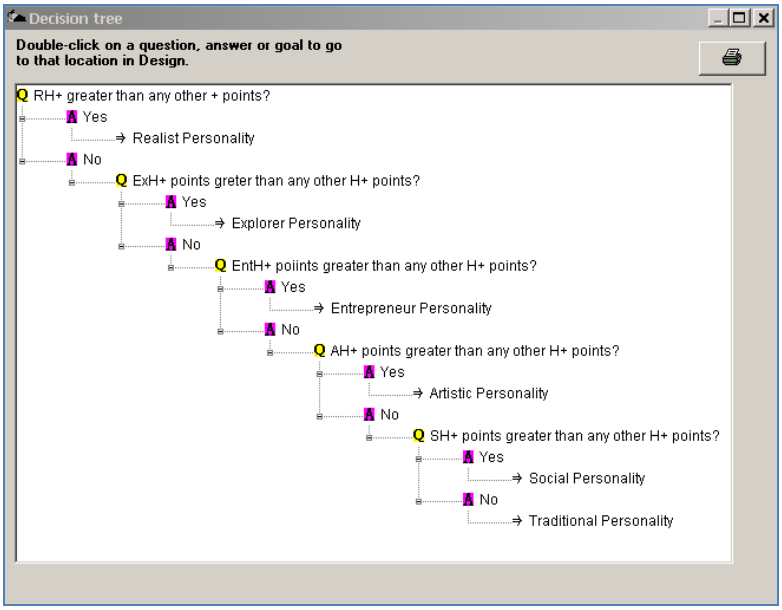

Figure 1. Decision tree of personality type determination.
with overall personality test scores

3.2. Incorporating Socio-Economic Factors Affecting School Selection into The Expert System Application

To incorporate socioeconomic factors into the model, expert sociologists' and pedagogues' support and factors discussed in the second section of the study are added to the model. 8 different criteria are defined affecting a primary school student's high school type selection decision. The socio-economic factors measured with survey application are listed in Table 2 with abbreviations.

Table 2. Socioeconomic factors.

No	Socioe-conomic factors	Abbreviation
1	Gender	GEN
2	Educational background of mother	EBM
3	Educational background of father	EBF
4	Inhabitation place	IPL
5	Income statue of the family	ISF
6	The educational background of fellows	EFE
7	Social activities	SOC
8	Time spent for reading book and newspaper	RBN

A questionnaire with 18 questions is used in order to measure the effect of each factor. Description of items used in the survey is listed in Table 3. Items in the survey are determined by brainstorming considering mostly addressed criteria in the literature.

For sample size determination Equation (1) is used. For a large population, with normal approximation, Equation (1) can be used for sample size determination [19]. The survey is conducted with a to high school students at 7 different type of high schools in Ankara. Although respondents are selected randomly, the sample size represents whole population (all primary school graduates in Turkey which is greater than 10.000) with 10% error rate. N: population size (>10.000); n:sample size (61); P: observed effect rate of *x* in the population (%80); Q: 1-P (%20); Z_α : standard z value for α= 0.05 (1,96); d: error rate (%10). With %10 error rate for 1.000.000+ students the sample size is determined as 61. The survey is conducted to 70 students. 8 respondents' questionnaire results are not taken into consideration because of inconsistency.

Table 3. Description of items used in the survey.

Parts	Questions	Item Descriptions
A. Demographic Data and Socio-economic statue	1-9	1. Gender 2. School type (7 type of school used in the study) 3. Class (1,2,3 or 4) 4. Number of brothers/sisters 5. A question asking whether father and/or mother is alive 6. Educational background of father 7. Educational background of mother 8. Inhabitation place of the family (city center, town or village) 9.Monthly net income of the family
B. Tendencies	10-14	10. Primary school graduated 11. Whether any of brothers/sisters attend a high school upstate 12. Major the student wants to work in the future 13. A question asking whether the student observes the activities of certain job types such as doctor, teacher, civil servant etc. at schools, hospitals and public institutions 14. A question asking to rank the factors affecting school type selection decision from 1 (least) to 5 (most) (Family, friends, teachers, future anxiety, counseling agency)
C. Social activities	15-18	15. Social association membership 16. Involvement in theatre, cinema and sports activities 17. Frequency of reading newspaper 18. Frequency of reading books

$$n = \frac{N \cdot P \cdot Q \cdot z_{\alpha}^{2}}{(N-1) \cdot d^{2}} \quad (1)$$

In order to determine the mathematical influence of each criterion, the correlation coefficients are determined with Pearson correlation tests. The survey has been conducted in science high school, social sciences school, Anatolian high school, vocational school, fine arts high school, military high school and general high schools. The data gathered from the questionnaire are analyzed in SPSS 11.5 statistics software package [20].

The most familiar measure of dependence between two quantities is the Pearson product-moment correlation coefficient or Pearson's correlation. It is obtained by dividing the covariance of the two variables by the product of their standard deviations. The correlation coefficient $\rho_{x,y}$ between two random variables *x* and *y* with expected values μ_x and μ_y and standard deviations σ_x and σ_y is defined as in Equation (2):

$$\rho_{x,y} = corr(x,y) = \frac{\mathrm{cov}(x,y)}{\sigma_x \sigma_y} = \frac{E[(x-\mu_x)(y-\mu_y)]}{\sigma_x \sigma_y} \quad (2)$$

where *E* is the expected value operator, *cov* means covariance, and, *corr* a widely used alternative notation for Pearson's correlation. The Pearson correlation is defined only if both of the standard deviations are finite and both of them are nonzero. It is a corollary of the Cauchy–Schwarz inequality that the correlation cannot exceed 1 in absolute value. The correlation coefficient is symmetric: *corr(x,y) = corr(y,x)* [21].

Table 4. The correlation coefficients between dependent variable school type and independent variables ISF, EBM, EFM and SOC.

Dependent Variable	Independent Variable	Pearson Correlation Coefficient	p Value
School Type	ISF	0.864	0.004
	EBM	0.742	0.001
	EFM	0.701	0.000
	SOC	0.655	0.000

If *x* and *y* have a strong positive linear correlation, *r* is close to +1 which is called positive correlation and if *x* and *y* have a strong negative linear correlation, *r* is close to -1 which is called negative correlation. If there is no linear correlation or a weak linear correlation, *r* is close to 0. A value close to zero means that there is a random, nonlinear relationship between the two variables. A correlation greater than 0.8 is generally described as strong, whereas a correlation less than 0.5 is generally described as weak [22]. According to the results, other variables GEN, IPL, EFE, RBN correlation coefficients are close to zero and not correlated with the school type selection. Therefore they are not included in the model.

In order to incorporate the effect of socioeconomic factors a multiple regression model is used in the study. The two predominant uses of multiple regressions are for prediction and explanation and the methodological approach taken in the analyses depends upon the purpose of the estimation of the model [23]. The specific multiple linear regression model according to survey results is given in Equation (3).

$$y = \beta_0 + \beta_{ISF} \cdot ISF + \beta_{EBM} \cdot EBM + \beta_{EFM} \cdot EFM + \beta_{SOC} \cdot SOC + \varepsilon \quad (3)$$

y: Dependent variable (School type)

ε: The stochastic or random source of variation

β_i: The regression coefficients *(i=ISF, EBM, EFM, SOC)*

The test for significance of regression is an F-test that determines whether there is a linear relationship between the dependent variable y and any of the independent variables. The null and alternative hypotheses for this F test are given in Equation (4) and (5) respectively.

$$H_0: \beta_1 = \beta_2 = \ldots = \beta_k = 0 \quad (4)$$

$$H_A: \beta_j \neq 0 \text{ for at least 1 } j \quad (5)$$

Rejection of the null hypothesis implies that at least one of the independent variables has a significant relationship to the dependent variable in the presence of the other variables in the model. This F-test is simply the ratio of the mean square regression divided by the mean square residual [24]. Table 5 shows the F-test results.

Table 5. F test results.

Model	**Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**
Regression	1609.936	61	536.645	279.794	0.000
Residual	772.057	388	1.918		
Total	2381.993	391	-		

Results show that variance between school type and independent variable is significant with 0.000 alpha level. In other words, school type selection changes according to independent variables ISF, EFM, EBM and RBN. Therefore, the null hypothesis is rejected.

The rules for socioeconomic statue considering the correlation degree of each independent variable are concluded as below:

IF EBM and EFM are “high school or higher” AND ISF >=2500 AND SOC=1 or 2

THEN socioeconomic score: 1

IF EBM and EFM are “primary school” AND ISF <2500 AND SOC=1 or 2

THEN socioeconomic score: 2

IF EBM and EFM are “high school or higher” AND ISF <=1500 AND SOC=1 or 2

THEN socioeconomic score: 3

IF EBM and EFM are “high school or higher” AND ISF >=2500 AND SOC=3 or 4

THEN socioeconomic score: 4

3.3.The Expert System Design

For graduating GPA, National Placement Test Score, the socioeconomic statue of the parents and results of the personality test; the rules, decision trees and decision tables of the system are created with Clixpert 8.0 expert system shell [25]. With these rules and decision trees the

knowledge-base is created. The inference engine executes the rules and interprets them to the user interface. The user enters the necessary data to the user interface. Finally the user can see the result with accompanying explanations. The system architecture is summarized in Figure 2.

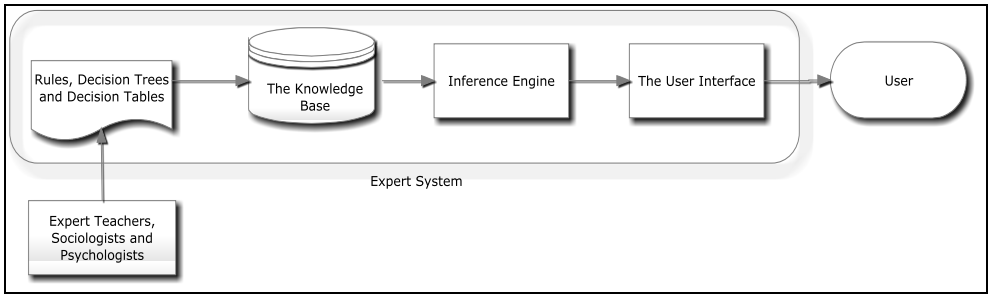

Figure 2. The system architecture.

After creating the knowledge base with academic achievement, social and socio-economic procedures the expert system is created with all of the rules and inference mechanisms. Basic mechanism of the expert system is summarized in Figure 3.

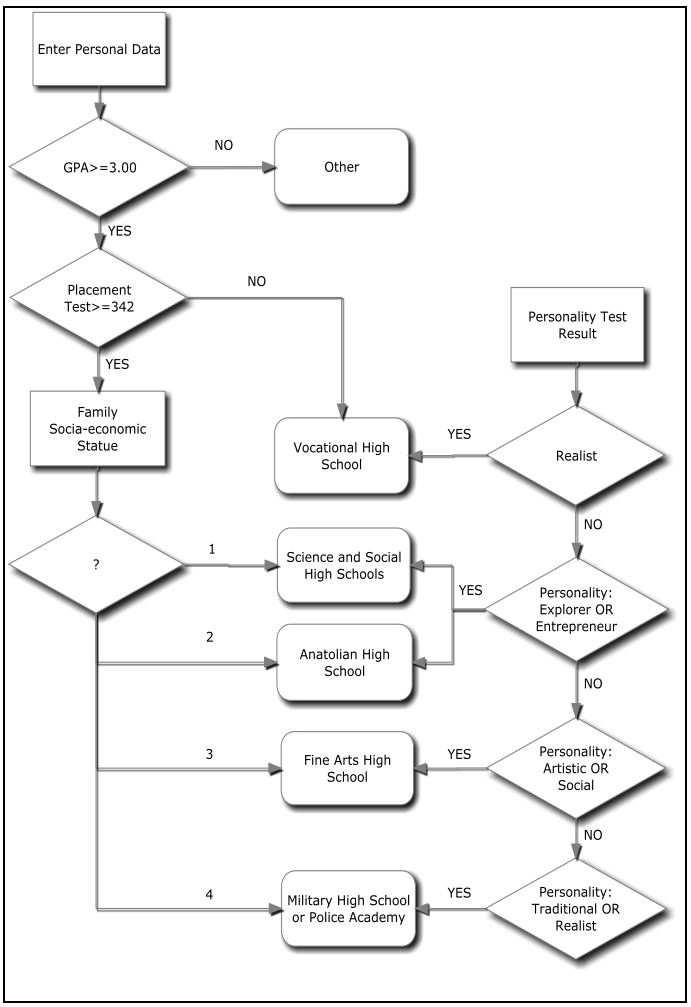

Figure 3. The basic algorithm of decision process.

The first important criterion for high school type selection is the graduating GPA score of the student. In the model GPA score of the student must be greater than 3.00 for being able select Science, Anatolian and Military

High Schools. The second criterion is the academic achievement in national placement test overall score which must be greater than 342.0 over 500.0 which is determined from Ministry of National Education of Turkey 2009 placement test base points data [26]. The third criterion determined is the socioeconomic status of the family. Science, Military and Fine Arts schools are scattered around Turkey. So the parents must be socially and economically available for covering the expenses when their children attend schools out of the resident. The last criterion is the personality test score. According to the experts explorer or entrepreneur personalities must choose Science High Schools, artistic or social personalities must choose Fine Arts high schools and traditional and realist personalities must choose military high schools. The knowledge-base for overall decision making process is created with Clixpert 8.0 expert system shell [25].

The user interface is developed with Visual Basic 2008 Express Edition [27]. Figure 4 shows the homepage of the user interface. Here the user enters the personal data to the program and takes mini-questionnaire and clicks the link for taking personality test. In Figure 5, the personality test screen is given. Here the user takes the personality test. For the purpose of ease of use the significant factors are included in the interface as a mini-questionnaire.

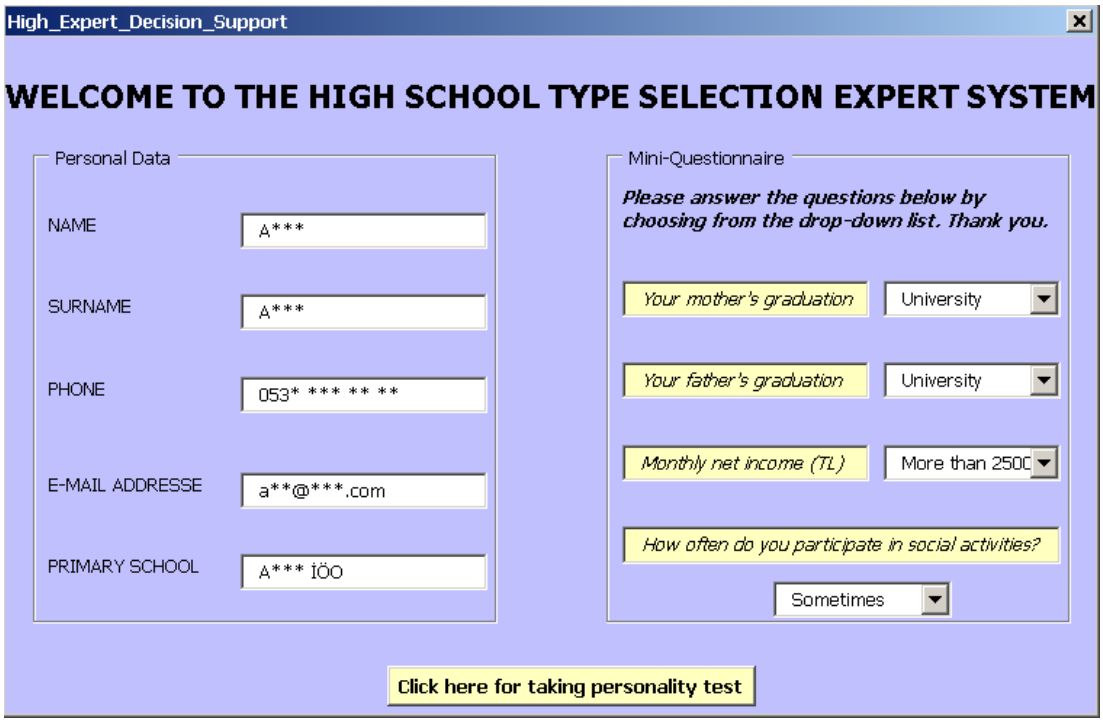

Figure 4. Homepage of the user interface.

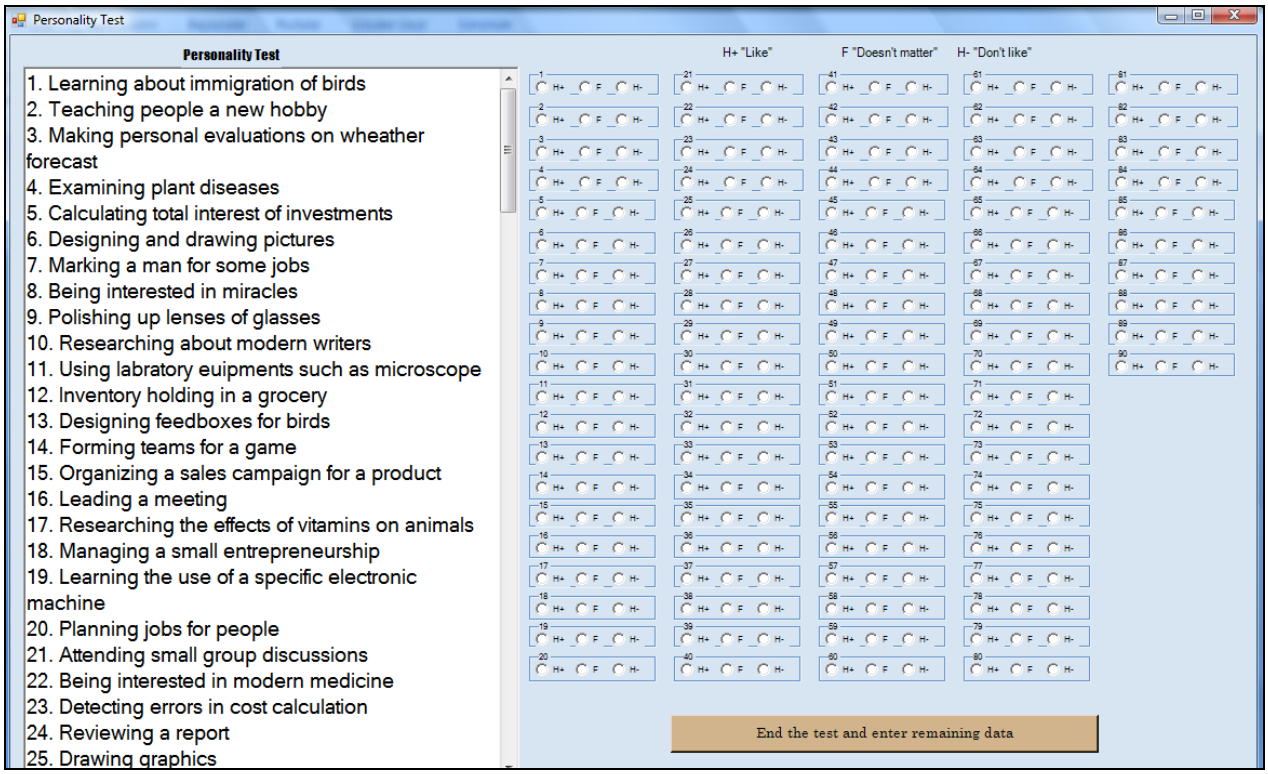

Figure 5. Personality test page.

In Figure 6 result page is seen. In the result page, the user can see the test result and enters placement test score, graduating GPA. Socio-economic statue score is calculated according to data entered on homepage of the user interface. Personality test result automatically comes to the screen. The socioeconomic statue score is automatically determined by the system according to the mini-questionnaire taken on main page of the software. Then the user can see the result in a different window together with explanations given in Figure 7.

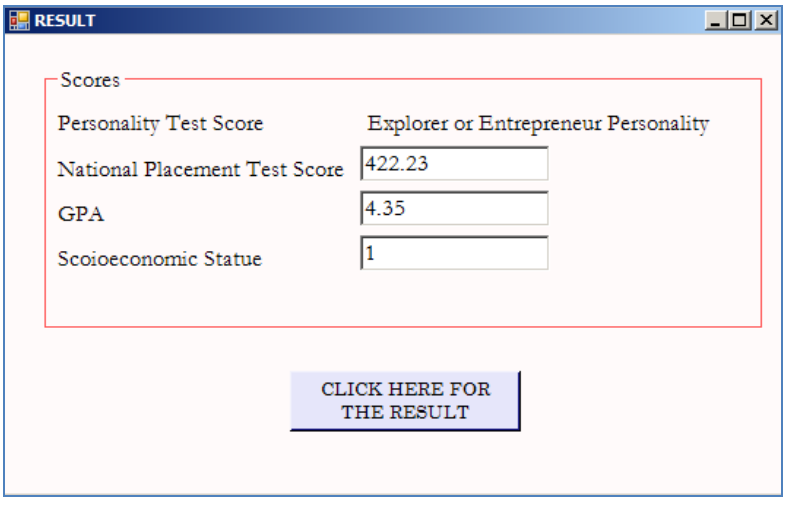

Figure 6. Result page.

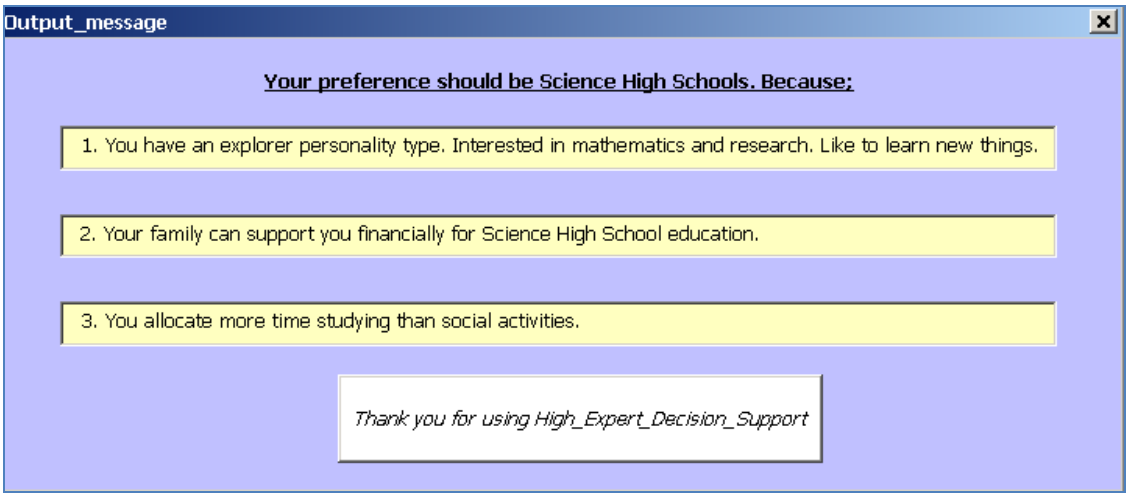

Figure 7. Output screen.

The applicability of the software is tested in terms of correctness and reliability for all possible scenarios with experts of a counseling center and with users namely students. The software works correctly to determine the seven type of school but it can be augmented by adding specific school types according to experts. This advice is considered as a future study. The software has advantage of wider distribution and access to scarce expert knowledge without consulting experts. In addition the system provides systematic and consistent application of knowledge. For similar or even for the same inputs, an expert may result in different decisions. However, the system always gives the same result for the same inputs.

4.CONCLUSION

In this study an expert system is designed, developed and implemented for high school type selection considering 4 important factors which are graduating GPA, national placement test score, the personality test score of the student and socioeconomic statue of the family. The expert system application created for this problem is a computer program that presents a recommendation to the student about high school selection. To integrate personality type variable to the model, specific rules are created. These rules are created according to information gathered from expert psychologists. ISF, EBM, EBF and SOC as socio-economic factors are added to model with rules which are created according to correlation degrees calculated according to survey results. Classical computer-assisted advisement programs' output is the list of schools considering only the centralized exam score. School selection as an important step of selection of profession should not be that simple. Other factors measuring tendency of the students must be taken into consideration in order to design a comprehensive system. For this reason, in this paper, a comprehensive model including technical and social criteria is created.

The system is designed for primary school graduates in Turkey. Similar applications may be encountered for specific education systems around the world. But for all of the students around the world, selection of profession should be determined after primary school. The study of Grupe (www.mymajors.com) [28], for instance, is an internet based expert system on major selection for high school graduates or collage freshmen. Students attending specific type of high schools such as social sciences or vocational high schools cannot choose every major. Therefore this type of systems should be directed at primary school graduates, before being late for trying to shape the future of an individual.

REFERENCES

[1] Reich, R.B., "The work of nations", ***Vintage Books***, New York, (1992).

[2] Aypay, A., "The tough selection at high school door: an investigation of the factors that lead students to general or vocational schools", ***International Journal of Educational Development***, 23: 517–527(2003).

[3] Kim, C.N., Yang, K.H., Kim, J., "Human decision-making behavior and modeling effects". ***Decision Support Systems***, 45: 517–527(2008).

[4] Benders, J., Manders, F., "Expert systems and organizational decision-making", ***Information & Management***, 25 (4): 207-213(1993).

[5] Archbald, D.A., "School selection and social stratification: Shortcomings of the stratification critique and recommendations for theory and research", ***Educational Policy***, 14 (2): 214–240 (2000).

[6] Eksi, H., Otrar, M., "Lise turleri ile kisilik ozellikleri arasındaki iliski uzerine bir arastirma", [A study on the relationship between high school types and personality], ***Kuram ve Uygulamada Egitim Bilimleri***, 1: 109–129(2001).

[7] Holland, J.L., "The Psychology of Vocational Selection A Theory of Personality Types and Environmental Models", ***New York Ginn. Pub.***, (1966).

[8] The Holland Tests. [Available from: http://www.hollandcodes.com/].

[9] Coleman, J., "Parent involvement in education". Policy Perspective, OERI: US Department of Education, Washington D.C. (1991).

[10] Epstein, J., "School-family relationships", ***Encyclopedia of Educational Research***, 6th Ed. Macmillan, 1139–1152 (1992).

[11] Lee, V.E., Croninger, R.C., Smith, J.B., "Parental selection of schools and social stratification in education: The paradox of Detroit", ***Educational Evaluation and Policy Analysis***, 164: 434–457(1994).

[12] Walker, M., Clark, G., "Parental selection and the rural primary school: Lifestyle, locality and loyalty", ***Journal of Rural Studies***, 26: 241-249(2010).

[13] Lai, F, Sadoulet, E, Janvry, A., "The adverse effects of parents' school selection errors on academic achievement: Evidence from the Beijing open enrollment program", ***Economics of Education Review***, 28:485-496 (2009).

[14] Nabiyev, V.V., "Yapay Zeka", ***Seçkin Yayınları***, (2003).

[15] Cawsey, A., "Lecture Notes" [Available from: http://www.macs.hw.ac.uk/~alison/ai3notes/].

[16] Liao, S.H., "Expert system methodologies and applications—a decade review from 1995 to 2004", ***Expert Systems with Applications***, 28 (1): 93-11(2005).

[17] Bo, W.K., Monaco, J.A., "Expert system applications in business: A review and analysis of the literature (1977–1993)". ***Information and Management***, 29 (3): 141-10 (1995).

[18] Bo, W.K., Monaco John A. "A bibliography of expert system applications for business (1984–1992)", 85-2 (7): 416-17(1995).

[19] Özdamar, K., "Modern Bilimsel Araştırma Yöntemleri", ***Kaan Kitabevi***, Eskişehir, 116-118(2003).

[20] SPPS version 11.5. [Available from: www.spss.com]

[21] Rodgers, J.L., Nicewander, W.A., "Thirteen ways to look at the correlation coefficient", ***The American Statistician***, 42: 59–66 (1988).

[22] Das, P., Bhattacharyya, D., Bandyopadhyay, S.K., Kim, T., "Person identification through IRIS recognition", ***International Journal of Security and its Applications***, 3(1): 129-147(2009).

[23] Pedhazur, E.J., "Multiple regression in behavioral research: Explanation and prediction", New York: Holt, ***Rinehart & Winston***, (1982)

.

[24] Smart, J.C., "Higher education: Handbook of theory and research", Vol. 17: Dordrecht, The Netherlands: ***Kluwer Academic Publishers***, (2002).

[25] Clixpert 8.0 expert system shell. [Available from: http://graemesummers.info/clixpert.html].

[26] [Available from: http://oges.meb.gov.tr/].

[27] Microsoft Visual Basic 2008 Express Edition. [Available from: http://msdn.microsoft.com/en-us/vbasic/default.aspx].

[28] Grupe, F.H., "An Internet based expert system for selecting an academic major", www.MyMajors.com. ***The Internet and Higher Education***, (5-4): 333-344(2002).