Classification and Profiles of Students Based on Their Motivations Concerning Higher Education

Anja Habus-Korbar[1], Natasa Tepic[1], Benjamin Culig[2], Vesna Luzar-Stiffler[3]
[1]*National Center for External Evaluation of Education (NCEEE), Zagreb*
[2] *Faculty of Humanities and Social Sciences, Sociology Department, University of Zagreb*
[3]*University Computing Centre – Srce and CAIR Center, Zagreb, Croatia*
anja.korbar@ncvvo.hr, natasa.tepic@ncvvo.hr, bculig@ffzg.hr, vluzar@srce.hr

Abstract. *Cluster analysis was used to create segments of high school graduates based on their motivations/opinions regarding their future education. Data for this study were collected using a questionnaire distributed to high school graduates following State Matura exam. The students were asked to evaluate the importance of 13 different reasons/causes for choosing the faculty selected as their first choice. The analysis yielded five meaningful clusters of students that differ not only in motivations but also in achievement on State Matura exams and in their preference regarding the scientific field of their future study.*

Keywords. motivation, students, university, State Matura, NCEEE, education, cluster analysis.

1. Introduction

During school year 2009/2010, a research project entitled "Student's Motivations toward General and Higher Education" was launched by National Center for External Evaluation of Education (NCEEE). The aim of this project was to assess the relationship between performance on State Matura exams and student motivations and attitudes concerning the adequacy of their high school (and their future higher) education. State Matura is a high stakes test. A high-stakes test is a test with important consequences for the test taker [1]. Since for the high school graduate it is the basis of a major decision, we may infer that achievement on the State Matura could be a measure of students' achievement/ success/attainment.

The first State Matura exam in Croatia was held at the end of academic year 2009/2010 by NCEEE [2]. Besides mandatory and elective exams, students were asked to complete a questionnaire. The questionnaire was voluntary and was used in this research as an instrument for assessing the motivation, etc. of high school graduates.

The questionnaire consisted of six sections/parts where students were asked to assess/answer a number of different issues: satisfaction/contentment with knowledge acquired in high school, preparation methods/techniques used for State Matura exams, assessment of their socio-economic status, and evaluation of the importance of 13 different reasons/causes for choosing the faculty graduates selected as their first choice.

In Croatia, State Matura scores, together with high school final grades, directly determine admission to a university, so applying for State Matura and university is a linked/related process.

2. Methodology

During the application process, students could choose ten different faculties and were obliged to rank them by preference. In this paper emphasis is on the last section where the students' assessed the importance of 13 different reasons/causes for choosing a faculty selected as their first choice.

2.1. Data Collection Instrument

On the instrument, the question was stated as follows: Please, on a scale of 1 to 5 (1 meaning not at all important and 5 meaning extremely important) rank the importance of each of the below listed reasons for choosing the faculty that you selected as your first priority. The offered reasons for selecting the faculty were the following:

- This profession will give me a greater chance of finding a job quickly after graduation (p31)
- This profession will provide me with a higher wage (p32)

- I always wanted to study this (p33)
- It's my parents' profession (p34)
- My parents persuaded me (p35)
- Its' a faculty my friends will apply to (p36)
- Its' a faculty with less applicants (p37)
- Some teachers motivated me (p38)
- It's a prestigious faculty (p39)
- It's not a challenging faculty (p40)
- This faculty is less expensive (p41)
- Applying to this faculty opens up possibilities for studying abroad (p42)
- It is easier to get scholarship for this faculty (p43)

For faculty and State Matura, 29245 students applied during academic year 2009/2010 of whom 29083 (99%) turned in their voluntary questionnaires, and approximately 23500 students provided answers to the block of questions used for this research.

Approximately 5500 students neglected to complete the second page of the questionnaire (because no instruction was provided on the instrument to "turn the page"); however no statistically significant differences were found between responders and nonresponders using test scores, gender or program (gymnasium, vocational).

2.2. Clustering Algorithm

The aim of this paper is to extract meaningful clusters of students based on their motivations/ opinions concerning enrollment to the university.

For this purpose clustering algorithm provided by SAS Enterprise Miner 6.1. was applied.

Clustering algorithm used is based on two methods – Ward's hierarchical and k-means partitional, and consists of three steps [3].

In the first step, 50 clusters were created using k-means method. Hierarchical Ward's technique was used in the next step to further cluster previously extracted clusters into yet a smaller number of clusters. The number of clusters was determined on the basis of Sarle's cubic clustering criterion (CCC) [4]. In the third step, the students were clustered again, using k-means method, with the number of clusters k provided in the previous step.

For a more detailed description of the characteristics of the clusters, we used segment profiling tool in SAS Enterprise Miner) [5, 6].

2.3. Cluster Stability

The stability of proposed cluster solution was checked using the following four step procedure:

1. A random sample of size 5000 was taken from the original data set of 24000 student responses.
2. Clustering algorithm described above (using k=5 clusters) was applied to the random sample.
3. The association between the original clusters and clusters obtained in step 2 was measured using chi-square and Cramer's V statistics.
4. Steps 1-3 were repeated 50 times, and the distributions of the two statistics were examined.

3. Results and discussion

As a result of clustering methods applied, five meaningful, homogeneous clusters were extracted.

The importance and distribution of each variable describing each of five clusters in relation to the distribution of the whole population is shown in Figures 1a, 1b, 2a, and 2b.

In Figures 1a and 1b we can see that the distributions on most important variables for cluster 1 (p42, p39, p43, p40) are left skewed

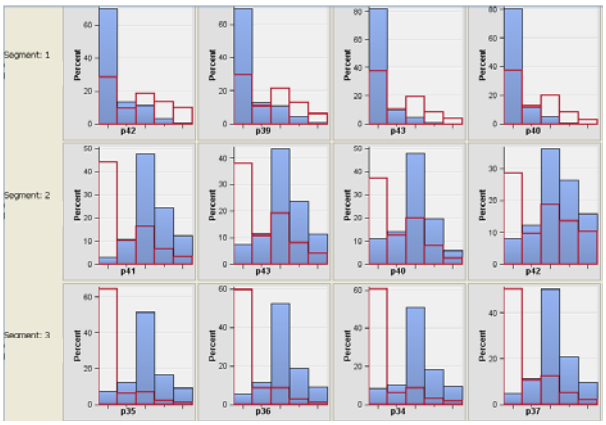

Figure 1a. Cluster 1-3 profiles

(filled bars) whereas the population distribution (empty bars) is more or less normal over the range from 2 to 5, with a peak at "1". In other words, students in cluster 1 found the reasons listed in questions p42, p39, p43, and p40 much less motivating than what the overall population did. The distribution of their answers is closer to the population distribution on answers p32 and

p31. The number of students in this cluster is 6605 (28%).

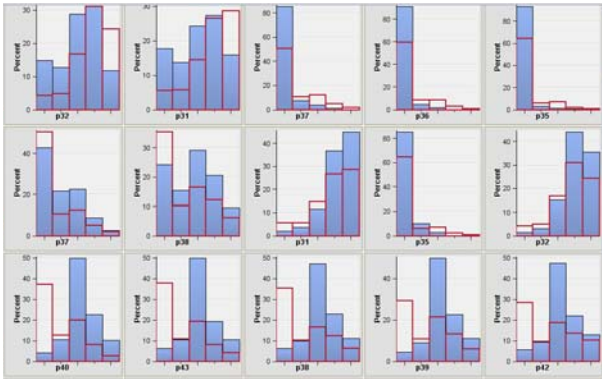

Figure 1b. Cluster 1-3 profiles (cont.)

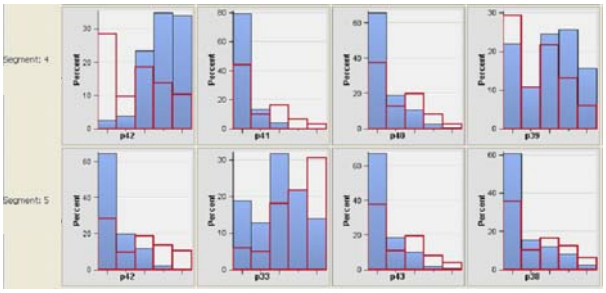

Figure 2a. Cluster 4-5 profiles

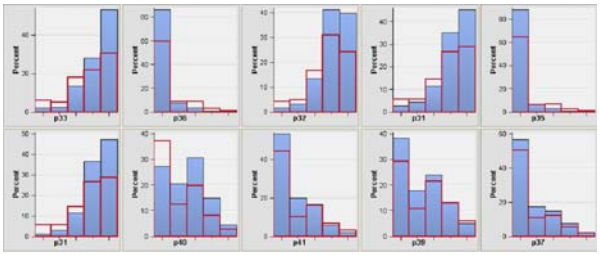

Figure 2b. Cluster 4-5 profiles (cont.)

The second cluster contains 4323 (18%) observations and is characterized by students giving unusually high importance to questions p41, p43, p40, p42.

Approximately 2600 graduates (11%) in cluster three selected questions p35, p36, p34, p37 as being critical for their decision regarding future education.

The distributions for clusters 4 and 5 are shown in Figures 2a and 2b.

Cluster four is characterized by more students giving higher importance to questions p32, p33, p31 and less to questions p40 and p41. The number of students in this cluster is 5440 (23%).

4580 graduates (20%) in cluster five gave higher importance only to questions p31 and p32, with all other questions (mostly p42, p33, p43, p38) being less important.

After summarizing the results, we concluded that graduates can be classified into five homogeneous groups based on their motivations. The groups are graphically displayed and sorted by the number of graduates in each (Figure 5).

First, and the largest cluster, containing 28% of all students are the ones we named “Not motivated” because of their indifferent motivation for everything except a stabile job and decent wage (prestigious faculty, studying abroad, scholarship, challenge is not interesting to them).

The “Scientists” (23%, cluster 4) in the next largest cluster, on the other hand, are the ones who want a prestigious, challenging faculty and are interested in studying abroad. They are also interested in finding a good position with good wage.

Cluster five (20%) are the graduates interested solely in money and good position (“Materialists”).

Students in cluster two (18%) we named “Easy-going”. They want a non-challenging faculty, would like to receive a scholarship and are concerned about the fees.

In the last, smallest, cluster number three (11%) we identified graduates we call “Pragmatics”. They are not really sure where to apply for study, are not interested in popular or challenging faculties; thus, their parents and friends have a considerable influence on their decision.

After understanding/explaining clusters using students’ motivations and opinions, an attempt was made to relate cluster membership to other relevant variables such as gender, preferred study area, and school program (gymnasium vs vocational) using correspondence analysis (please see Figure 6). Along the first dimension are students with preferences towards technical/ biotechnical fields distinguished from artistically oriented, those who chose biomedical/medical, humanities, social and natural sciences. Gymnasium students, “Scientists” are positioned on upper side of dimension two, while students from vocational schools, “Easy-going”, “Not motivated” and “Materialists” are found at the lower side of the second dimension. There appear to be four groups of students:

- “Pragmatics”, mostly male, with preferences towards technical areas;

- "Scientists", coming from gymnasiums favoring either natural sciences, humanities or biomedical/medicine;
- "Not motivated" and "Easy-going", mostly female students with interest either in art or social studies; and
- "Materialists", primarily coming from vocational schools wishing to apply to biotechnical faculties.

Cluster stability was checked using the procedure described in Section 2.3 above. Monte Carlo distributions of chi-square and Cramer's V statistics for 50 random samples are presented in Figures 3 and 4.

The results show that the chi-square values (for the association between the original clusters and clusters obtained from random samples) have a mean value of 9717 and standard deviation of 2078, and that their distribution is approximately normal. All chi-square values (with 16 degrees of freedom) are highly statistically significant.

Distribution of Cramer's V statistic (which measures the strength of association) is also close to normal, with mean of 0.69 and standard deviation of 0.07. Values for all 50 random samples demonstrate high association between the original clusters and clusters obtained from random samples, which confirms that the original clusters are relatively stable to data perturbations.

4. Conclusions

After identifying groups of students by their motivations it can be concluded that great majority of high school graduates are interested in finding a job with a good wage. The students that achieved best results on State Matura exam choose prestigious faculties and show a lot of interest in continuing their education abroad. The faculties they select as their first priority are in the field of natural sciences, biomedicine, medicine and humanities.

We believe that the observed cluster groups will provide valuable information to all parties involved (schools, universities, Ministries, etc.). The findings of this research will be used in future studies concerning education in Croatian high schools and for continuous monitoring and improvement of Croatian education system.

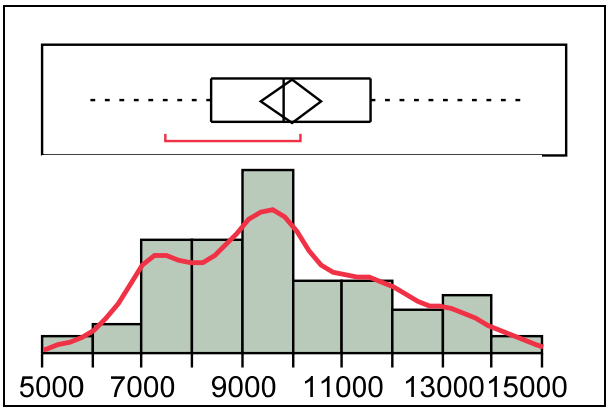

Figure 3. Monte Carlo distribution of chi-square statistic

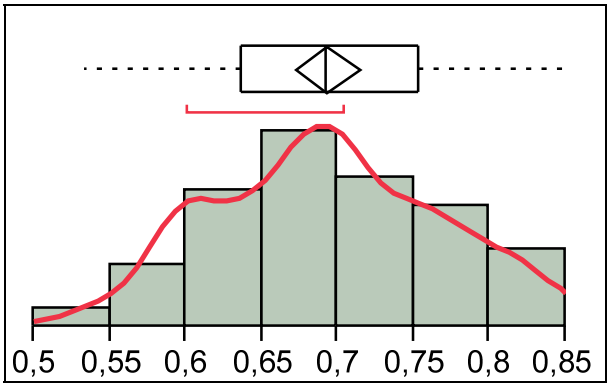

Figure 4. Monte Carlo distribution of Cramer's V statistic

5. References

[1] A Lexicon of Learning. http://www.ascd.org/Publications/Lexicon-of-Learning/Lexicon-of-Learning.aspx

[2] State Matura and national assessment reports; 2007-2010. http://www.ncvvo.hr [01/31/2011]

[3] Jain A.K., R.D. Dubes. Algorithms for Clustering Data. NJ. Prentice Hall Inc., 1988

[4] Sarle W.S. Cubic Clustering Criterion. SAS Technical Report A-108. Cary, NC: SAS Institute. 1983,

[5] Hand D.H. Mannila, P. Smyth. Principles of Data Mining. Cambridge, Massachusests, The MIT Press. 2001.

[6] SAS Institute Course Notes. Applied Analytics using SAS Enterprise Miner 5.3, Cary, NC: SAS Institute. 2008.

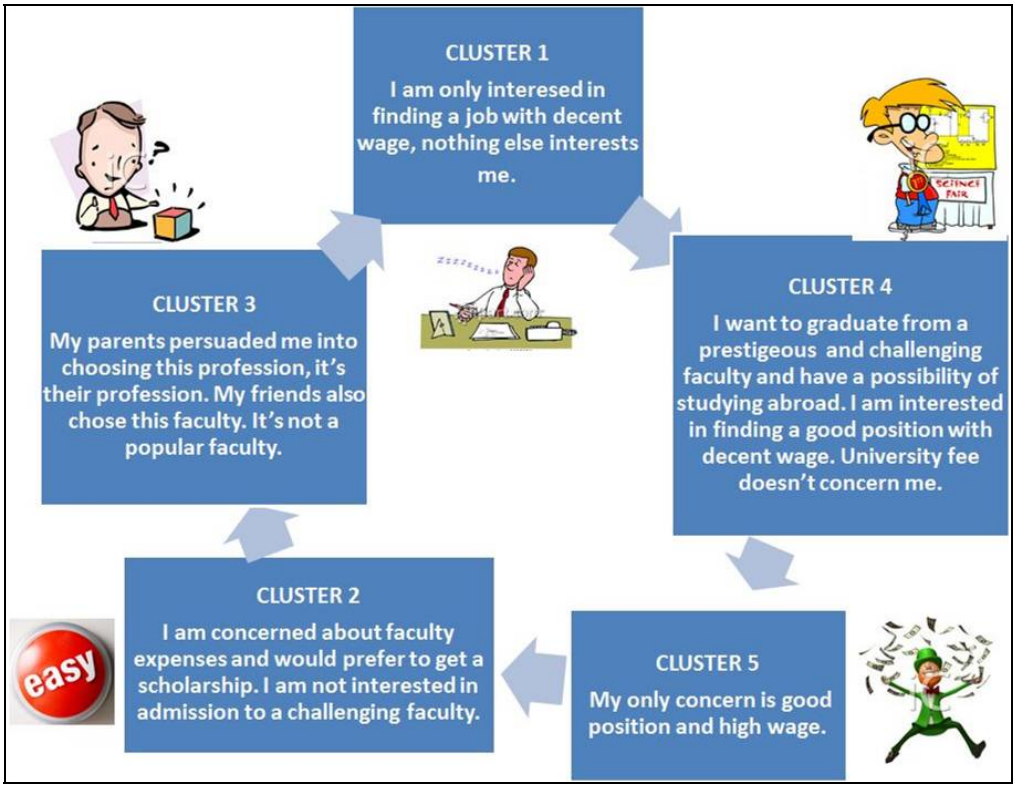

Figure 5. Cluster description

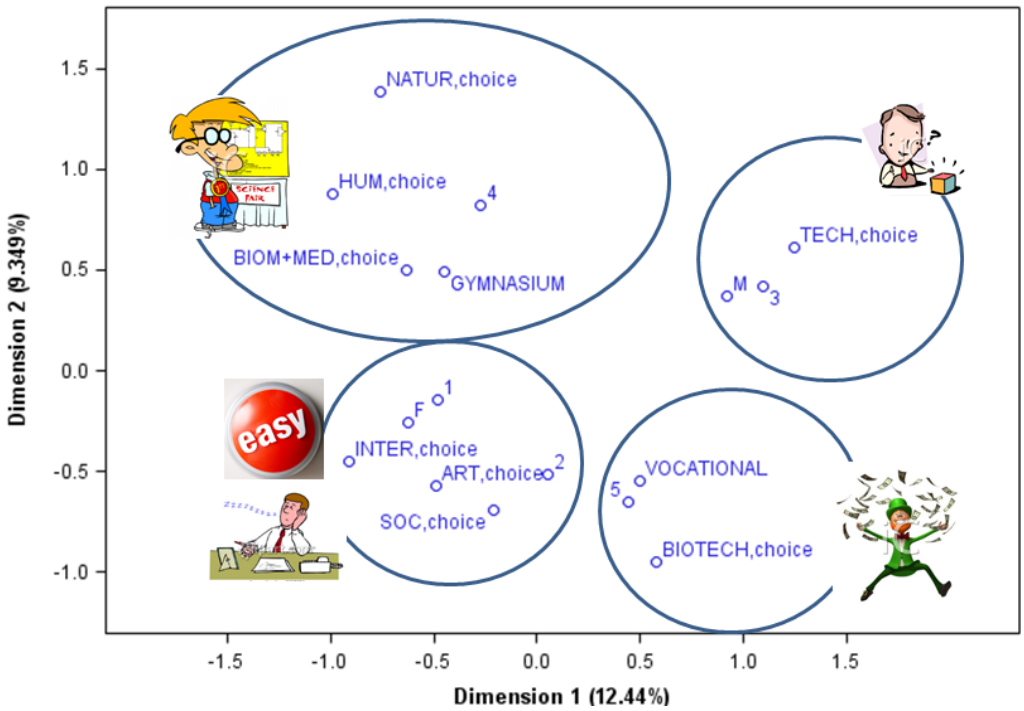

Figure 6. Correspondence analysis plot for association among clusters (1-5), gender (M,F), program (gymnasium, vocational), study area preferrence (Natural Sciences and Math, Humanities, Biomedical/Medicine/Vet, Technical, Biotechnical, Social studies, Art, Interdisciplinary)