CLASSIFICATION OF INDIVIDUALS ACCORDING TO THEIR OPINIONS ON ACQUIRING NECESSARY COMPETENCIES WITHIN THEIR STUDIES

Marta Žambochová1; Aneta Kulhanová2
Jan Evangelista Purkyně University in Ústí nad Labem,
Faculty of Social and Economic Studies, Department of Mathematics and Computer Science,
Moskevská 54, 400 96 Ústí nad Labem, Czech Republic
e-mail: 1 marta.zambochova@ujep.cz; 2 aneta.kulhanova@seznam.cz

Abstrakt
New times require new skills in new areas. This is the reason why the structure and syllabus are becoming obsolete. Many teams of teachers study new content of teaching materials and new ways of teaching. These efforts are supported by different organizations, beginning with the leadership of schools across the state authorities to the bodies of the European Union. In our study, students’ point of view is examined. The group of respondents consists of both potential and current students, as well as the former ones. The emphasis is laid on finding skills which "ordinary people", according to their opinion, need for their personal and professional life. The aim of the research was to determine the knowledge and skills that should be part of the studies according to these people, and on the other hand, to point out things which they would gladly give up. The survey was conducted among schools at all levels, although the priority was the college level studies. Our expectations were confirmed in many ways. However, many of the requirements were inconsistent with the actual state of education. The knowledge gained in our research could help in creating track objects, their content and the form of teaching.

Keywords
Professional competence; Undergraduate students; Exploration; Tests of hypothesis; Decision trees.

Introduction
New time requires new knowledge in new fields. That is the reason why the structure and syllabi of teaching become obsolete. The problem with improving necessary competences within teaching at different school levels has been solved by professionals from many countries (e.g. Russia, Belgium, Spain, USA, Iran, the Netherlands, or Cyprus).

It is very important to adapt to increasing specialization and differentiation of knowledge and to use proper method of teaching, especially at universities [15]. Therefore, the methodology of appropriate professional competences has to be explored [5], [11], [22]. Nowadays, the problem with IT competences is relevant. Several conceptual frameworks for the development of professional competences based on their own empiric research were presented [1].

Moreover, rethinking the field of foreign languages education is also current [16]. The criticism of changes in syllabi and concepts of approach to pupils in the field of language education is present [7]. Through the research among students it was concluded that the realized changes had not brought expected improvement, particularly concerning communication abilities. According to the opinion of respondents, schools wrongly primarily focus on improvement of grammar, whereas the communication competence is neglected.
The research among Japanese students at American universities showed that the greatest shortage is in the field of communication and interactive and presentation competences [16]. There is a demand for strengthening the role of experience and practical knowledge during foreign language studies as well as emphasis on communication abilities [10]. Finding new approaches, methods and means of creation of foreign language communicative competences of future university foreign language teachers is important [6]. Except for IT and foreign language fields, there are also other fields of coveted education, e.g. competence in natural science [14], [17] or entrepreneurial education [3], [9], [12].

1 Research Objectives

Research was repeatedly performed within student activities at the Faculty of Social and Economic Studies of Jan Evangelista Purkyně University in Ústí nad Labem between the years 2012 – 2014. The first phase of processing is mentioned in article [28]. By using a questionnaire survey, this research monitored opinions on the level of information acquired at school, on demands, and on reality. The research focused on people residing in the Ústí Region of the Czech Republic. The respondents were people over 18 years old who finished their studies no more than 15 years before their participation in the research. Within processing this research, respondents were classified from different points of view, namely by the age, gender, level of education, the field of study and also by their opinions on the quality of education and acquiring different kinds of competences. Two basic types of classification were used – a cluster analysis and classification trees. The objective of this classification was to analyse subjective respondents’ view of competences that school system provided them. The objective was also to analyse respondents’ view of competences that were, according to their opinion, important during discharge of their occupation. The focus was particularly put on the structure of these competences and their level.

2 Data

The questionnaire was answered by 954 respondents older than 15 years. 181 of these questionnaires were answered in 2012, 284 in 2013, and 489 in 2014. Men represented 47% of respondents. The average age of respondents was 33.

Their education structure is apparent in Fig. 1. It is obvious that secondary school graduates predominated in the research. The structure of respondents in terms of the number of previous employments is apparent in Fig. 2. The education structure is apparent in Fig. 3. It is obvious that business school graduates predominated in the research. The second most frequent field is the technical one. It is obvious in Fig. 4 that employees predominated in the research in terms
of current job focus. Half of all addressed employees were represented by administrative employees. Almost a quarter of respondents were represented by students.

It was also found out that 74% of respondents expressed satisfaction with their field of education, 35% of respondents are studying at university at the moment (either full-time or part-time apart from being employed), and 38% of respondents plan further education.

3 Methods

The questionnaire mostly consisted of closed questions. That means that respondents could choose from provided answers. They were often answers assigned to a five-grade scale. Only a minority of questions was open questions that were to specify the answers from previous closed questions. According to the character of questions quantities of different types were formed, nominal categorical quantities, ordinal quantities as well as cardinal continuous ones.

While processing the data, different kinds of statistical hypothesis testing were used. Diverse groups of respondents were compared from different points of view. In the case of categorical non-scale quantities the chi-square test for independence was used. In the case of ordinal variables (at scale questions) the Friedman test was used to assess the level of answers in different groups. In the case of continuous quantities a pair of tests was used: the \( F \)-test of the equality of two variances and then the Student’s \( t \)-test to compare mean values (based on the result of the \( F \)-test, the \( t \)-test was determined with comparable variances, in the case of the \( t \)-test with incomparable variances). See the usage description in [18]. All tests have been done at a 5% level of significance. The null hypothesis was the equality of observed quantities (variances, population means) of individual sets.

One of the main steps of our processing was the classification of respondents in terms of their opinions with regard to education quality within the meaning of acquiring knowledge necessary for being successful at work. For this purpose, two basic classification methods were used: the cluster analysis and decision trees. The work uses the method of the questionnaire-based surveys and data processing with the use of relevant survey methods described, for example, by [8] or [2].

3.1 Cluster Analysis

The cluster analysis is one of the methods of education without a teacher [4], [20]. It deals with similarity of data subjects. It addresses division of the group of subjects into several previously non-specified groups (clusters) so that the subjects within individual clusters are as
most similar as possible and subjects from different clusters are as least similar as possible. This happens via minimization of the selected purpose function.

The cluster analysis may be performed by many different methods. Individual methods differ by various methods of determining similarity of subjects (similarity rates) and also by clustering methods (e.g. hierarchical and non-hierarchical). While selecting the cluster analysis method it depends on whether the source data or aggregated data is directly available (e.g. the table of frequencies or similarity matrix). If the source data are available, the selection method depends on the type of variables (nominal, ordinal, quantitative variables). In order to process our data, the selection of the hierarchical method was not suitable because of a relatively large number of subjects. The \( k \)-means algorithm is intended for clustering subjects described by quantitative variables, which was not our case. In order to use this method, it would be necessary to pre-process the data by binarization, which means by a transfer of each categorical variable to several binary variables (a variable acquiring only 0 and 1 values). The most suitable method for processing the data was the two-step method.

The principles of the two-step cluster analysis are described for example in [19]. This method uses the BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies) algorithm which is described in detail in [26], or [27]. The algorithm creates a so-called CF-tree to which continuously incoming data are assigned. The advantage of this procedure is that it goes through the data set only once. The disadvantage is sensitivity to the order of incoming data points.

3.2 Decision Trees

Decision trees belong to the group of methods of education with a teacher. In this case the decision-making rules for assigning subjects to classes are created on the basis of the teaching (training) group.

Various types of decision trees are a widely used group of trees utilized in data models. Decision trees are structures that recursively sort out observed data according to certain decision-making criteria. The tree root represents the entire population set. The internal nodes of the tree represent subsets of the population set. The values of the explained variable can be read in tree leaves. Two types of decision trees are used, classification trees (each leaf has a class assigned) and regress trees (each leaf has a constant assigned – estimate of the explained variable value).

The decision tree is created recursively by splitting the space of predictor values (explaining variables) on the basis of finding a question (branching condition) that splits the space of observed data into subsets as well as possible, that means it maximizes the criterion of splitting quality (so-called splitting criterion).

The splitting process stops if the so-called stopping rule is met. Another step of the algorithm is tree pruning. It is necessary to determine the “proper” tree size (too small trees do not sufficiently characterize all relations within data; too big trees include also arbitrary data properties in the description). Sub-trees of the tree that are created by the construction algorithm are generated and the quality of sub-trees generalization is compared (how well they depict data).

The procedure may be that decision trees are initially created based on so-called training data and then their quality is validated on so-called testing data. Another method is cross validation when all data are used to create a tree and its sub-trees. Afterwards the data are divided into several disjunctive parts of a similar size and gradually one data part is removed from the set. By using the formed data sets the quality of a tree and its sub-trees is validated. Such a sub-
tree which has the lowest estimate of an actual error is selected. If more sub-trees with a comparable estimate of an actual error exist, the smallest sub-tree is selected.

A huge amount of algorithms was developed in order to create decision trees. The most used are CART, ID3, C4.5, AID, CHAID, and QUEST. Three types were used in our work. Their algorithms are implemented in the SPSS statistical system.

The CART algorithm (Classification and Regression trees) [21], [25] can be used in the case where there is one or more explaining variables available. These variables may be either continuous or categorical (ordinal and nominal). Furthermore, in our work a variable which may be also categorical (nominal and ordinal) or continuous was once explained.

The CHAID algorithm (Chi-squared Automatic Interaction Detector) [25] is a modification of the AID method for a categorical dependent variable. The result is non-binary trees. For testing, the method uses the chi-square test.

The QUEST algorithm [13] is usable only for the nominal dependent variable. Similarly to the CART algorithm, only binary trees are created. Unlike the CART method, which performs the selection of the variable for splitting the node and selection of the splitting point during the tree creation at the same time, the QUEST method performs these tasks separately. The QUEST (for Quick, Unbiased, Efficient, Statistical Tree) method eliminates several disadvantages of algorithms using the exhaustive finding (e.g. CART), such as time expenditure of processing, decrease in result generality.

4 Results and Discussion

While comparing the data sets from different years, no statistically important difference in individual sets was found in most cases (the resulting $p$-values were considerably higher than 5%, they were within a range of 0.26 to 0.74). Then, it was possible to continue in working with all sets in summary. The only exception was the issue of university currently studied at. This question had a wide range of answers and the composition in individual years was very different. For this reason, this question was not included in the main processing.

One of our main research questions was the importance of acquiring competences in IT and foreign language fields. Therefore the focus of this research is on these subjects in more detail in several detailed questions. After performing the statistical evaluation the conclusion was that improving abilities in these two fields would be presumed by the majority of people (almost 95%) even beyond school education. The tests proved that this opinion was shared by all groups of respondents. Only the $p$-value 0.12 pointed out slight difference of unemployed people. This category of people prefers extracurricular improvement in 98%.

However, there was a contradiction found in requirements for the type of competences in these fields. The respondents were satisfied with the extent of hours allotment to subjects related to IT and foreign languages but, contrarily, they were dissatisfied with the amount and kind of knowledge acquired. This fact was most evident at university education (p-value for IT was 0.032, for foreign languages it was 0.028). By a closer research it was found that just these respondents feel insufficient connection of school and practice. Therefore, it is possible to conclude that these subjects are taught at school rather theoretically and less practically than the respondents would need in their further employment.

Another part of questions was focused on improving competences that do not directly belong to specific subjects. An example of these competences is an ability to learn, logic, stress resistance, reliability, adaptation to conditions, working with people, self-presentation, etc. Within the entire investigated set, surprisingly logic had the highest statistical difference ($p$ value 0.017) the improvement of which would be required by 85% of people across education.
levels and fields. Another considerable group was the competence of working with people, reliability, ability to learn and adapt to specific conditions and self-presentation ($p$ value within $0.028 – 0.039$). All these competences were supported by approximately 60% of the respondents. At some of these competences, statistical tests showed differences within individual groups of respondents. The ability to adapt is, for example, strongly supported by people without the secondary school-leaving exam. The ability to work with people is considerably supported by administrative employees. People of artistic specialization and executives would like to develop the ability to self-present.

As already stated above, one of our main steps of our processing was the classification of the respondents in terms of their opinions with regard to education quality within the meaning of acquiring knowledge necessary for being successful at work.

In the first classification stage, the cluster analysis was carried out in order to divide the respondents into groups with similar opinions. As a classification variable a triad was chosen, describing the rate of satisfaction with the school studied at, with the rate of willingness to be educated in IT and foreign languages even outside school. The two-step method divided the respondents into five groups with quite a good quality of resulting clustering. This is apparent in Fig. 5.

![Cluster Quality](image)

**Source:** Own research

**Fig. 5:** Cluster quality of model with 3 inputs and 5 clusters

Furthermore, three algorithms were used to create decision trees, namely CRT, CHAID, and QUEST. The aim of these trees was to find the description of created clusters in terms of the observed respondents’ opinions. The second objective was to find the description of created clusters in terms of the respondents’ identification. In both cases the respondent’s affiliation to individual clusters was chosen as the explained variable. In the first case, the explained variables were the triad of variables identical to classifying variables of the cluster analysis. In the second case, the explained variables were identification data of the respondents, namely gender, age, education degree and field, current employment, and data on current or future education.

The quality of the second group of decision trees was very high. Risk estimate was within 0.080 to 0.177. This means that the successfulness of classification was within 82.3% to 92.0%. In this case, the tree created by using the QUEST algorithm was of the highest quality. Unfortunately, the quality of the first group of decision trees was rather low. Risk estimate was within 0.502 to 0.556. This means that the successfulness of classification was at low level within 44.4% to 49.8%.

Despite this, it is possible to come to several interesting conclusions about our respondents. Entrepreneurs, executives, and non-administrative employees are more or less satisfied with the level of the school they studied at in terms of the acquired knowledge and they are strongly willing to be educated within IT even outside school education. Students of secondary general education and bachelors of humanistic and law studies are not satisfied at
all with the level of the school they studied at in terms of the acquired knowledge but they are not very willing to be educated outside school, not even within IT or foreign languages. Graduates from lower school education of pedagogical, medical or IT fields are considerably satisfied with the level of the school they studied at and they are willing to be further educated within IT and foreign languages. Graduates from secondary professional schools, economy and IT bachelor studies are strongly willing to be further educated within IT.

In the second classification stage, the cluster analysis was carried out in order to divide the respondents into groups with similar opinions on different studied subjects. As the classifying variables, there were twenty variables selected, expressing detailed respondents’ opinions on teaching these selected subjects (among them for example mother tongue, maths, foreign languages, PC work, psychology, law basics, financial literacy but also general overview). The two-step method divided the respondents into two groups. The quality of the resulting clustering was somehow lower than in the previous case. Yet it is still on the usability threshold.

Furthermore, three algorithms were used again to create decision trees, namely CRT, CHAID, and QUEST. The aim of creating these trees was again to find the description of the created clusters. At first, the aim was to find the description of the created clusters in terms of the respondents’ identification. The second description was in terms of observed respondents’ opinions. In both cases, the respondents’ affiliation to one of the two resulting clusters was chosen as the explained variable. In the first case, the explained variables were identification data of the respondents, namely gender, age, education degree and field, current employment, and data on current or future education. In the second case, there were chosen twenty variables identical to classifying variables of the cluster analysis as the explained variables.

The quality of the group of decision trees was balanced and very high. Risk estimate was within 0.088 to 0.116. This means that the successfulness of classification was within 88.4% to 91.2%. In this case, the tree created by using the CRT algorithm was of the highest quality. This tree is shown in Fig. 6. The quality of the second group of decision trees was higher than in the case of the previous clustering. Risk estimate was within 0.275 to 0.354. This means that the successfulness of classification was at well-usable level, exactly within 64.6% to 72.6%.

The most distinct results of this classification can be summarized into the following conclusions. Graduates of lower schools of humanistic, pedagogical, nature-science studies and schools oriented on IT, who graduated at the turn of the century, have the opinion that mother tongue and PC work should be taught for the whole duration of education. They consider the ability to adapt to external conditions to be an important property that should be supported by the education process. Young respondents of the age up to 24 years who graduated at general, economic and artistic schools of a secondary or bachelor level belong to a group that would prefer to be educated in foreign languages for the whole duration of their studies. On the contrary, they do not consider mother tongue education for the whole duration of their studies to be beneficial. They do not have any clear opinion on teaching exact subjects and IT.

In the third, the last stage of processing, selected respondents’ answers to individual questions regarding their satisfaction were chosen as the classifying variables. The two-step method created three clusters. The information about the respondents’ affiliation to the cluster in the processed variables was included. The quality of the resulting clustering was very high.
Fig. 6: Classification tree created using the algorithm CRT
In another step of processing, all three types of classification trees were created again. The affiliation to the cluster was chosen as the explained variable; and the answers to individual questions regarding the respondent’s opinions as explaining variables. Risk estimate is ranged from 0.142 to 0.235 for all the created trees. This means that the success rate of classification of objects ranged from 85.8% to 76.5%. The quality of models was adequate. Furthermore, the second triad of decision trees was created. The affiliation to a cluster as the explained variable was chosen again. Unlike in the previous processing, the answers to questions regarding the type of respondent as the explaining variables were chosen. Risk estimate is ranged from 0.348 to 0.483 for all the created trees. This means that the success rate of classification of objects ranged from 65.2% to 51.7%. Therefore, the quality of models was not very high. But this quality can be considered sufficient.

The result of our classification is the following findings. The most satisfied respondent is over 35 years old, of economic and technical specialization. These respondents mostly miss education within foreign languages and ICT. People with pedagogical specialization are less satisfied but they have no specific wishes. On the contrary, the least satisfied ones are the respondents with secondary economic education working in administration, of the age between 23 and 35. They mostly miss higher connection of school and practice as well as higher quality of teachers. Apart from foreign languages and ICT they would prefer more education in the field of entrepreneurial competences, logic but also psychology and general overview. They would also prefer education in the form of discussion and assistance with improving self-presentation ability. The respondents of artistic specialization have no considerable objections to the quality of the school system. They would prefer more creativity. It would be even preferred by medical students. Apart from this, they would prefer better connection to practice and improvements within working with people.

Conclusion

Our research shows that, despite rumours, people are generally more or less satisfied with the quality of the school system in the Czech Republic. This issue is most critically perceived by people of secondary economic education who work in administration. The biggest problem of the Czech school system is seen in the connection of school and practice. This confirms the views of the authors of the article [23]. The example of the development model of competencies which graduates need in the labour market is also described in the article [24]. There is also dissatisfaction with the competencies of teachers.

Obviously, there is a demand for more intensive foreign language and ICT education across all respondent groups. These subjects should be taught for the whole duration of education. In this field, there is also dissatisfaction with the quality of teachers and teaching methods. In both fields the respondents would prefer less theory and more practical experience. This conclusion is in compliance with conclusions of several international studies, such as [7], [16], [11], or [6]. Unfortunately, our research also implies that not all those who perceive insufficient competences in these fields also have a need to be further educated outside the school system itself. This mainly applies to graduates from secondary general education and humanistic and law bachelor studies.

However, the view of other education orientation is more differentiated according to respondent groups. Technically oriented respondents perceive insufficient quality of maths education in the Czech school system. Humanistic- and economic-oriented respondents would prefer to have more psychology, logic, entrepreneurial education and general overview. Respondents with medical and artistic specialization would like to get higher support for creativity. Respondents of the two latter groups would prefer education in the field of working with people. It is therefore obvious that higher differentiation of subjects at different school
specializations is necessary, preferably in the form of various optional subjects. Younger respondents, unlike other ones, do not perceive the need to improve their competence to adapt to external conditions. The question is why. It is possible that young people already (or yet) do not consider this competence to be so important.

An interesting phenomenon worth pointing out is how the opinion on the given issue differs between men and women. There appeared to be no statistically considerable differences in any of the tests. This was also confirmed after the inclusion of the “gender” item in the classification methods. None of the used methods assesses gender as an important identification. It is therefore apparent that men and women have similar opinions, which is satisfying.

Our conclusions are more or less in compliance with investigations from other countries mentioned at the beginning of this article. It is apparent that the mentioned problems of the Czech school system are not only the national issue but one of an international character.

Acknowledgements

The research is financed by the internal grant agency of the J. E. Purkyně University in Ústí nad Labem in the Czech Republic.

References


KLASSIFIKACE JEDNOTLIVCŮ PODLE JEJICH NÁZORŮ NA ZÍSKÁNÍ POTŘEBNÝCH KOMPETENCEÍ V RÁMCI JEJICH STUDIÁ

Moderní doba vytváří stále větší tlak na individuální vzdělávání člověka a tím kladě stále větší důraz na co nejefektivnější získávání vědomostí. Pro poskytovatele těchto informací je důležité znát potřebu skladby požadovaných kompetencí a způsobů předávání znalostí. Tvorbu nových osnov vyučované látky i novými přístupy k výuce se zabývá mnoho týmů pedagogů, a to na různých úrovních počínaje vedením škol, přes státní úřady až po orgány Evropské unie. V naši studii jsme se zabývali subjektivním pohledem studentů, a to jak současných, tak i bývalých. Zaměřili jsme se především na zkoumání schopností, které by běžný absolvent školy potřeboval dle svého názoru pro svůj osobní i profesní život. Šlo nám o analýzu kompetencí, které by dle oslovených respondentů měly být součástí studia, a které jim naopak připadají nadbytečné. Průzkum byl proveden pro různé úrovně škol, i když naší prioritou byla především vysokoškolská úroveň studia. Srovnáním s jinými studiemi jsme zjistili, že situace v České republice není příliš odlišná od jiných zemí. Samozřejmě jsme nalezli i mnohá specifika pro Českou republiku. Získané poznatky by mohly pomoci při tvorbě osnov a formy výuky na různých úrovních školství.

KLASSIFIZIERUNG VON PERSONEN NACH IHRER MEINUNG ZUM ERWERB NOTWENDIGER KOMPETENZEN IM RAHMEN IHRER UNTERSUCHUNGEN


KLASYFIKACJA JEDNOSTEK WEDŁUG ICH OPINII NA TEMAT POZYSKANIA NIEZBĘDNYCH KOMPETENCJI W RAMACH ICH STUDIÓW