ANALYSIS OF EXAM RESULTS OF THE SUBJECT 'APPLIED MATHEMATICS FOR IT'

Helena Borožová, Jan Rydval

Czech University of Life Sciences Prague

Highlights

- *The exam results of "Applied Mathematics for Informatics" from the last 13 years are analysed*
- *The reliability, difficulty, and discrimination power of the tests are analysed*
- *Students' frame influences their examination results*

Abstract

In this paper the exam results of the subject "Applied Mathematics for Informatics" (formerly Methods of Operation Research) from the last 13 years have been analysed. The exam has two parts: written test and oral exam. The grades of the students of the subject have been low for a long time. We would like to know if the low grades are due to the quality of the tests or due to a reducing number of hours of contact teaching. Another reason of low grades can be due to the mathematical character of the subject and to the unpopularity of such kind of subjects. Based on the bad results, students have also initiated a change in the scoring system.

This article builds on our paper at the conference ERIE 2013. The main goals of this paper are to find out if the grades have had the tendency to decline during the years and to evaluate the validity, reliability, difficulty, and discrimination power of the tests.

Introduction

Each subject has its own preferences and expectations that create its own view (frame). As Tversky and Kahneman (1981) mentioned the framing effect is made up of these frames. They also pointed out that framing effect influence the way in which the information is interpreted. Individual decisions are influenced by the presented information and by the problems formulation (Druckman, 2001). Therefore, the framing effect can be defined as a set of preferences and expectations of involved subjects belonging to a particular decision-making problem. Specific frames can also be defined for the teachers´ and students´ point of views (Rydval and Brožová, 2011). These frames can negatively affect passing the information in the education process. Rydval and Brožová (2011) also mentioned that the usual student's frame aims to succeed in examinations with the least effort. This influences students' results significantly.

Examination, testing, test scoring and grading are very important parts of a pedagogical work. The purpose of these activities is to assess student's knowledge related to a subject. Student's grade for a course is generally based on a scoring on the final exam and the oral examination. It is necessary to have a suitable quality and validity of tests and a good objective scoring system to ensure consistence of the students grading.

Basic statistical analysis of test results is a method typically used in many study information systems or e-learning systems. On the other hand, these tools give only summary results, as frequency of the grades, average of the grades and number of attempts. However, more detailed analysis has to be done by teachers themselves. Kaspříková (2011, 2012a, 2012b) did such

analysis for instance for mathematics course at University of Economics in Prague and Jarkovská et al. (2012) for distance programs at Czech University of Life Sciences Prague (CULS Prague).

Jacobs (1991), Miller (2012) and Wells and Wollack (2003) analysed following other characteristics of the tests:

- Test validity – test should be suitable for the objective testing of the students from the three points of view:
  - Content validity – test should be able to assess student's knowledge of the subject.
  - Criterion validity – test should be able to measure student's knowledge.
  - Predictive validity – test should be able to predict, for example, student's knowledge during the oral exam.
- Test reliability – test should be reliable and consistent and should not be subjected to random errors.
- Test difficulty – test should not discourage students from further learning; it should be neither too difficult nor too easy.
- Test discrimination power – test should show a difference between skilled and unskilled students, the skilled students should answer test questions well.

The aim of this paper is to analyse the basic characteristics of the exam test in the subject of Applied Mathematics for Informatics (AMI), earlier Methods of Operation Research (MOR). This analysis is performed because the students consider the test as

very difficult and due to the students´ grades that have been too low.

This article follows the contribution of Brožová and Rydval (2013) in which the results of the subject "Applied Mathematics for Informatics" from the years 2009/10, 2010/11 and 2011/12 were analysed. The exam results of the last 13 years are analysed and also the impact of changes to the scoring system is evaluated.

The main questions which would have to be answered are:

- Have the results had the tendency to decline in the last 13 years?
- Has the test been very difficult?
- Which scoring system is more suitable?

Material and Methods

Content and structure of the exam tests

The subject Applied Mathematics for Informatics (AMI) is in the curriculum as a specialization of Informatics in regular and distance study programs at the Faculty of Economics and Management, Czech University of Life Sciences Prague. In the last 13 years the test results of these subjects have not reached the satisfactory level from the teachers' perspective, because majority of students (more than 70% since the year 2008/09) reached only grade 3 - good or didn't pass the exam. Therefore, in this research, the exam tests used in these subjects are analysed and their characteristics are discussed.

The main subject topics are covered by the lectures and seminars, definitions and steps of algorithms are highlighted during the teaching. During the last lecture the brief recapitulation of the subject content is made. Moreover, the structure of the test is described along with the scoring and grading system. The exam has two parts, written test and oral exam.

In total, more than 30 variants of the test exist, and they differ in selected topics and numerical input. These variants were used during the last 13 years with only small changes, which were always done according to the actual content of subject.

The test is scored and the total possible score is 100 points. The minimum amount of points necessary for the oral examination is 50 points. The grading system uses three grades: 60 – 73 points is a good grade, 74 – 86 points is a very good grade and 87 – 100 points is an excellent grade.

During the oral examination students have to confirm their knowledge related to the subject. Nevertheless, students can increase their score, i.e. improve their grade.

The test is divided into 3 parts. The first part consists of three theoretical questions, the second part includes two small examples and, finally, the third part consists of a large practical example. The questions and examples of the tests follow all the topics of the subject.

- Three theoretical questions – these short questions have a form of a brief question that requires a written answer not longer than a few sentences or a paragraph. For example the students have to describe or explain the basic definitions, the steps of algorithms or calculations or the simple principle. Maximum score of each question is 10 points.
- Two small examples – these questions have a form of computational questions, which have to be solved by more or less simple calculations and the results have to be interpreted. Maximum score of each example used to be 15 points. The scoring was changed last year and,

nowadays, maximum score of each example is 20 points.

- Practical example (essay) – this part of the test has a form of a case-study / scenario question, which is used to prove that students can understand and integrate key concepts of the course, apply theory to a practical context, and demonstrate the ability to analyse and evaluate obtained results. Depending on the problem description of the small practical problem the students have to select and create a suitable model, solve it, and interpret the results. Maximum score used to be 40 points. The scoring was changed last year and, nowadays, maximal score is 30 points.

To analyse the exam results we use the data from Student information system from the last 13 years from 2000/2001 to 2012/2013; we collected data such as number of students, their grades and number of attempts.

For the detailed analysis of exam tests, the scores of test items, the total test scores and the final grading have been collected from the last four years, regardless of the number of attempts. Together we have 325 tests from the year 2009/10, 265 tests from the year 2010/11, 193 tests from the year 2011/12, and 292 tests from the year 2012/13.

Methods used for analysis of the tests

The high quality exam tests help to evaluate the student's knowledge and motivate the students to learn. In this research, we use the following methods for analysis of the test quality (Jacobs, 1991; Miller, 2012; Wells and Wollack, 2003):

- Difficulty Index of the tests,
- Discrimination Index of the tests,
- Reliability of the tests.

The analysis of the tests is supplemented by an overview of the exam results. The following parameters are calculated:

- Average grade,
- Average number of exam attempts,
- Success rate.

Difficulty Index

Difficulty index (P) of the test questions is one of the most useful and the most frequently reported analyses. It is a measure of a proportion of examinees who answered the question correctly; for this reason this index is frequently called as P-value:

$$P = \frac{S_{sum}}{S_{max}} \qquad (1)$$

where S_{sum} is a total number of obtained scores of all students; S_{max} is a maximum possible amount of score.

Difficulty index can range between 0.0 and 1.0. The higher value indicates that a greater proportion of examinees responded to the question correctly, or in the other words the higher the value the easier the question is. The index of difficulty of a suitable question lies in the closed interval [20%, 80%] (Škoda et al., 2006).

Discrimination Index

Discrimination index (ULI - Upper-Lower Index) is a measure we use to distinguish between good and bad students (the students are ranked according to their scores).

$$ULI = \frac{N_U - N_L}{0.5N} \qquad (2)$$

where N_U is the number of good students, students from better group who answered the question properly;

N_L is the number of bad students, students from worse groups who answered the questions properly;

N is the total number of students.

The possible range of the discrimination index is -1.0 to 1.0; however, if a question has the discrimination index below 0.0, it suggests a problem. A negative discrimination index indicates that the question (test item) measures something other than the rest of the test.

The values of the discrimination index and the difficulty index have to be interpreted together, because there is a relationship between them. If an item has a very high (or very low) P-value, the potential value of the discrimination index will be much smaller than if the item has a mid-range P-value. The questions are suitable if the difficulty index is from [30%, 70%] and the discrimination index is greater than 0.25. If the difficulty index lies within the interval [20%, 30%] or [70%, 80%], the discrimination index has to be greater than 0.15 (Škoda et al., 2006).

Difficulty (P-value)	$[20\%, 30\%]$	$[30\%, 70\%]$	$[70\%, 80\%]$
Discrimination (ULI)	≥ 0.15	≥ 0.25	≥ 0.15

Table 1 Recommended values of difficulty and discrimination indices

Reliability of the tests

Measure of reliability can be calculated using the method for measurement of the internal consistency (Cronbach, 2004; Škaloudová, 2012). Reliability in this way shows if all test items' content is homogeneous, if these items measure the same knowledge with the similar score. The Cronbach's alpha evaluates the test items using multi-scale scoring for reliability calculating

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum_{i=1}^{k} s_i^2}{s^2}\right) \qquad (3)$$

where s_i^2 is the variance of the *i*-th test items score;

s^2 is the variance of the test score and

k is number of test items.

The value of the α coefficient of reliability varies from 0.0 (no consistency) to 1.0 (perfect consistency). The coefficient α is only a lower bound of the reliability, so the real reliability is often much underestimated. Obviously, a larger coefficient is better (Cronbach, 2004; Revelle and Zinbarg, 2009).

Cronbach's alpha	Internal consistency
$[0.9, 1]$	Excellent
$[0.8, 0.9]$	Good
$[0.7, 0.8]$	Acceptable
$[0.6, 0.7]$	Questionable
$[0.5, 0.6]$	Poor
$[0, 0.5]$	Unacceptable

Table 2 Meaning of values of Cronbach's alpha

It is acceptable for subject's exams to have lower reliabilities because the grades are based on several measurements – at least on written test and oral examination, and also each student can take the exam three times in the worst case (Wells and Wollack, 2003; Jacobs, 1991).

Exam results analysis

Average grade – average students' grade is calculated only for results of the students who passed the exam, it means for grades 1 - excellent, 2 – very good and 3 – good, as a sum of a collection of grades divided by the number of successful students.

Average number of attempts – average number of attempts is calculated as a sum of all used exam terms divided by a number of studying students.

Success rate – Success rate is calculated as a ratio of number of the successful students and number of all students.

Results

As it was mentioned above the main questions which need to be answered are:

- Have the results had the tendency to decline in the last 13 years?
- Has the test been very difficult?
- Which scoring system is more suitable?

Analysis of the exam results

Data about the exam of the subject Applied Mathematics for IT as the number of students, their grades and the number of the exam attempts were collected from the Student information system from the year 2000/2001 to 2012/2013. For the main characteristics of students' results see Figure 1, Figure 2, Figure 3 and Appendix for a Table 3.

It is possible to say that average grades for group of regular students were slightly increasing so the grades show the tendency to be worse. In accordance with this fact, the success rate was decreasing and the number of attempts was increasing. This is demonstrated by the logarithmic trend lines. Logarithmic trend was selected, because the analysed values have upper or lower bound. Coefficient of determination is greater than 0.6. Year 2011/2012 was the year in which the old subject and the new subject were taught together and the students of the old subject had to study very hard, because repetition of the old subject was no longer possible. This is the reason for partial improvement of the results. Development of the distance students' results is different, because these students have different reasons for their education. These students need a university education for their jobs and so they are forced to study.
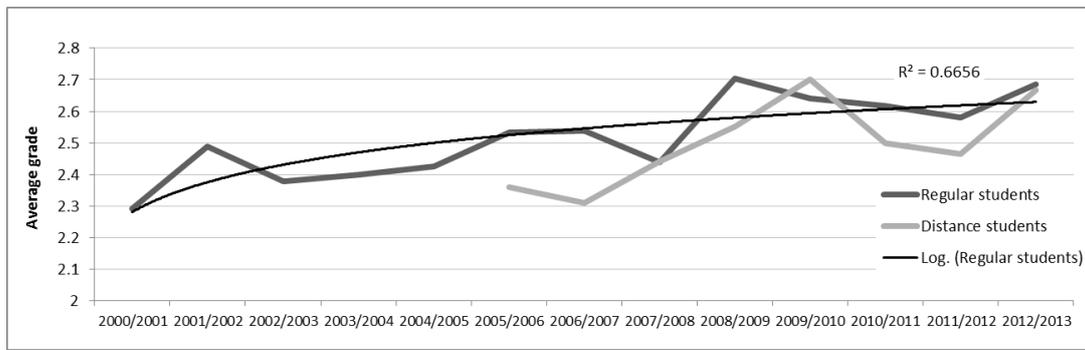
Figure 1 Development of average grade, 2000/01 to 2012/13 (source: own calculation)
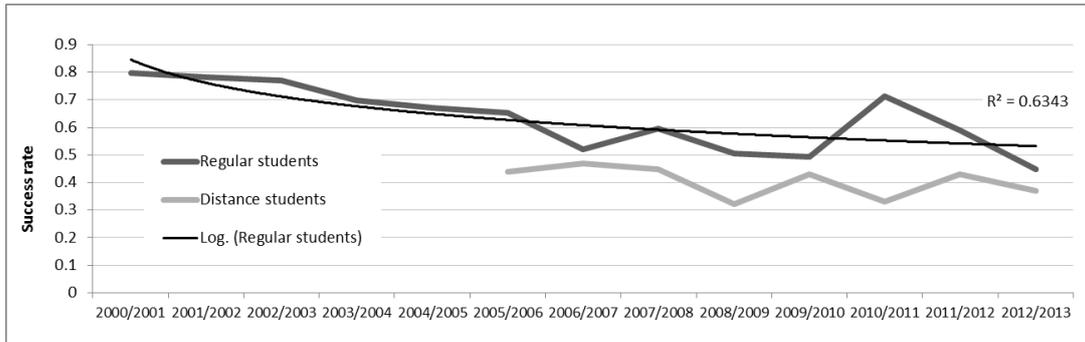

Figure 2 Development of success rate, 2000/01 to 2012/13 (source: own calculation)
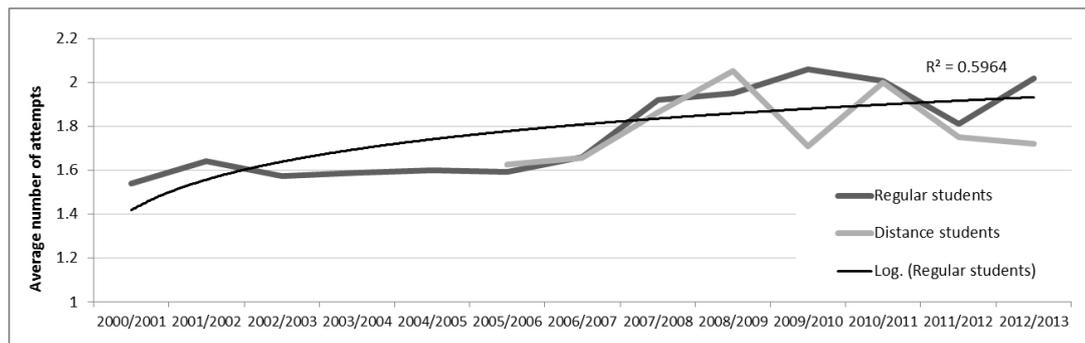

Figure 3 Development of average number of attempts, 2000/01 to 2012/13 (source: own calculation)

Analysis of the tests quality

Data of the tests scoring were collected from the year 2009/10 to 2012/13. The tests are scored from 0 to 100 points. The frequency of the number of the points is calculated for the unequal intervals (Figure 4 and see Appendix for Table 4), because at least 50 points are necessary for the oral exam and from 60 to 73 points is necessary for 3 – good, 73 to 86 for the grade 2 – very good and 87 to 100 for the higher grade 1 – excellent. It is surprising that about 50% of the tests are scored less than 50 and more, about 30% of the tests are scored less than 30. Students with such test did not pass the exam, so it is possible to suppose that many students come to the exam to try it and to find out what the exam tests are like. That reason students also explain during the oral examination. However this strategy means that they lose one exam attempt. This faithfully corresponds with the usual student's frame which is to succeed in examinations with the least effort.

Discrimination index (ULI) and the Difficulty index (P-value) of the tests were calculated for the whole test and also for each item of the test (see Appendix for a Table 5).

The column '*P-values ALL*' contains the Difficulty index for all tests in the group. Difficulty index in the column '*P-values >50*' was calculated for the group of tests with the whole score higher

than 50 points. The answer scored at least with 60% of points is considered as a correct answer.
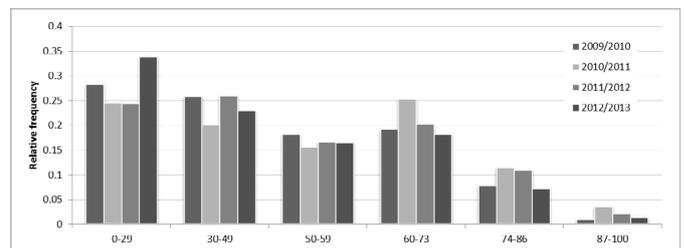

Figure 4 Frequency of the number of the test points (source: own calculation)

Cronbach's alpha was calculated for the whole test and also for questions and examples of the test (see Appendix for a Table 5). The column '*Cronbach's alpha ALL*' contains the Reliability index for all tests in the group. The Reliability index in the column '*Cronbach's alpha Second half*' was calculated for the second half of the tests in the exam session.

Difficulty index values are between 0.35 and 0.56 for all tests from different years. This index is higher for the group of the tests scored more than 50 points (between 0.44 and 1.00). This can be explained by the fact that the students were better

prepared for the resits. Difficulty index calculated for all tests has a satisfying value; therefore the tests have good levels of difficulty (Figure 5 and Figure 6).
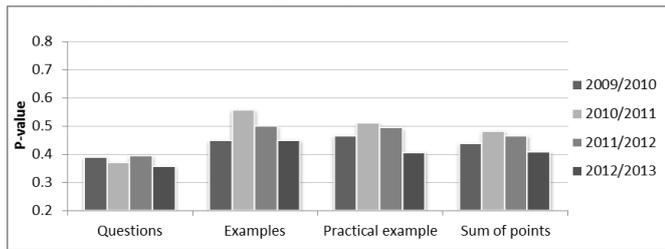

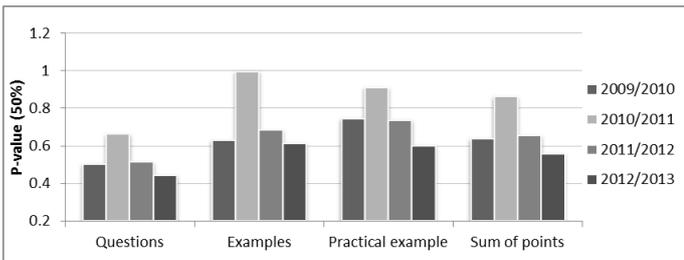**Figure 5 Difficulty index for all tests in the group (source: own calculation)**

Figure 6 Difficulty index for tests with a whole score higher than 50 points (source: own calculation)

Values of the Discrimination index of all tests are greater than 0.5, so the tests distinguish well between good and bad students. The Discrimination index of the theoretical questions is the lowest, it seems, that the students seek to know practical application of studied method and not the theoretical background (Figure 7).
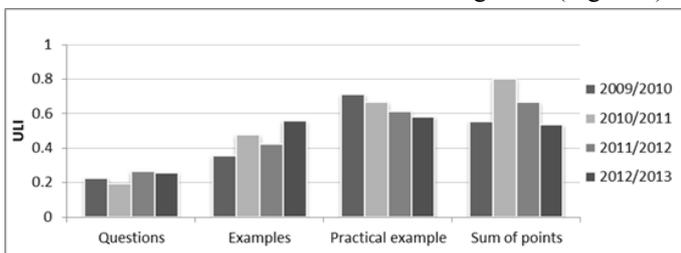

Figure 7 Discrimination index of the tests (source: own calculation)

The Cronbach's alpha of the whole tests lies in all cases within the interval [0.5, 0.659]. It is not a high reliability because the reliable tests have the Cronbach's alpha near to 1.0. However, because the students have to make the resits, if they do not pass the exam, the conditions are different; the students learn more during the second or the third attempt and, therefore, reliability values of the second half of the tests in the exam session are better (Figure 8 and Figure 9).
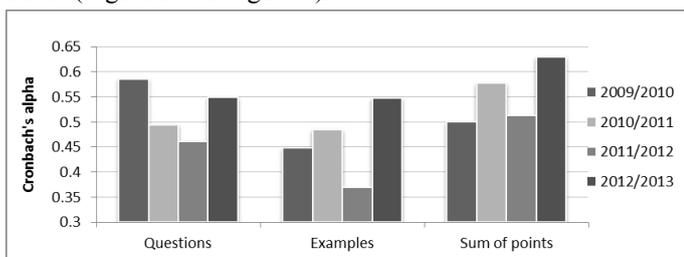

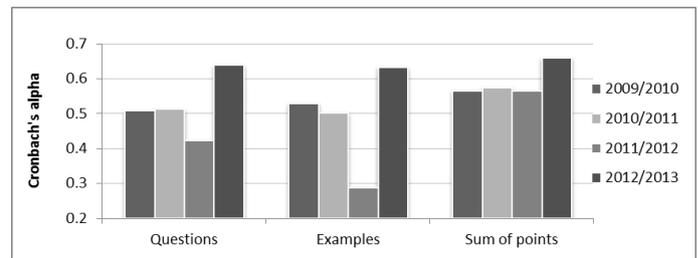**Figure 8 Cronbach's alpha of all tests (source: own calculation)**

Figure 9 Cronbach's alpha of the test from the second half of the exam session (source: own calculation)

In the year 2012/2013 the scoring system was changed, 10 point for the correct answer of theoretical question remained, but the small examples are awarded by 20 point instead of 15 and the practical example by 30 instead of 40 points. The reliability of the test was increasing (Figure 8, Figure 9 and see Appendix for a Table 5) to 0.659 with this new scoring system, so the decision to change the scoring system was good. This value of the Cronbach's alpha is still not satisfactory. Nevertheless, it is necessary to consider a system where students can three times repeat the exam and, therefore, the tests are formulated in 30 variants and also each part of the test is aimed to determine the different types of knowledge - definitions, calculations, and practical applications.

Discussion

The decreasing tendency of the average grades for the group of regular students and the success rate together with the increasing number of exam attempts show the worsening of the exam results. This may be caused due to the mathematical character of the subject and unpopularity of such kind of subjects, the reduction of the number of hours of seminars since 2011 and the students' frame of the least effort.

Nowadays, universities recognise that students are entering higher education system with a poor mathematical preparation and lower level of basic mathematical skills (Gallimore and Steward, 2014; Grossman, 2001). Therefore, the lack of sufficient mathematical knowledge can affect students' achievements on Operations Research courses. In addition, Jordan et al. (1997) report that 77% of instructors view the mathematical background of students or fear from mathematics as a principal source of teaching and learning problems.

Teaching students a mathematically-oriented course with insufficient mathematics backgrounds has predictable results: frustrated instructors, frustrated students, and poor teaching ratings. Operations research and computer science use tools of mathematics to solve and analyse problems. Students who lack a lucid appreciation for mathematics are limited in their ability to understand and explore the Operations Research and Computer Science interface (Hardin et al, 2012).

The difficulty of the tests were increasing (P-value was decreasing) although the same tests were used repeatedly from the year 2000 (Figure 5, Figure 6 and see Appendix for a Table 5). This fact can also be caused by the reduction of the number of hours of the seminars; formerly each topic has been planned for a 90-minute long lecture and a 90-minute long seminar. However, from the academic year 2011/2012 only 45 minute long seminars are planned.

The worsening tendency of the test results may not be caused only by the reduction of hours of the seminars, but could partially correspond with the traditional students´ effort to go through a learning process using the path of least resistance. This traditional student´s decision frame means the student is

satisfied with the least result and adjusts study effort only for passing an exam. Not always the students realise that the path of least resistance is not the most satisfactory for the future. Such students' frame is confirmed by several time-used studies in engineering education, which show that students use less time studying than was allocated in the curricula (Kollari et al., 2008).

Conclusion

Analyses of 13 years series of the grades of both regular and distance students show a slight increase in the difficulty of the tests and, therefore, together with the students' frame of the least effort the grades have had the tendency to decline. However, in the last four years this trend has been slowing or stopping (Brožová and Rydval, 2013).

Very disturbing is the very high number of the tests with less than 50 points, this fact apparently shows that students use the first exam term to only become familiar with a form of the test and the exam. However, the information about the form of the test is provided during the last lecture and, therefore, students waste their exam terms.

Analysis of the scoring system shows that the new scoring system 10-10-10-20-20-30 is preferable, because the results are not so dependent on a practical example. Test reliability increased, but the value is not satisfactory, which is primarily due to a possibility of the two resits and also due to the small number of the test items. Test reliability (Cronbach's alpha is greater than 0.5 for whole tests) can be considered satisfying because we include also resit tests and each test consists of only 6 items.

The tests have appropriate difficulty (P-values are between 0.4 and 0.5 for whole tests). For students it is the hardest to answer the theoretical questions. So we have to pay more attention to the careful construction of the test questions. We have to phrase each question clearly so students know exactly what they are asked for. The discrimination power of the tests is high (ULI values are greater than 0.5 for whole tests) which means that the test structure and used questions are suitable.

Acknowledgement

References

Brožová, H. and Rydval, J. (2013) 'Analysis of the exam test quality', *Proceedings of the 10th International Conference on Efficiency and Responsibility in Education,* Prague, pp. 120–127.

Cronbach, L. J. (2004) 'My Current Thoughts on Coefficient Alpha and Successor Procedures', *Educational and Psychological Measurement,* vol. 64, no. 3, pp. 391-418. http://dx.doi.org/10.1177/0013164404266386.

Druckman, J. N. (2001) 'Evaluating framing effects', *Journal of Economic Psychology*, vol. 22, no. 1, pp. 91-101. http://dx.doi.org/10.1016/S0167-4870(00)00032-5

Gallimore, M., Stewart, J. (2014) 'Increasing the impact of mathematics support on aiding student transition in higher education', *Teaching Mathematics and Its Applications*, vol. 33, no. 2, pp. 98-109, http://dx.doi.org/10.1093/teamat/hru008

Grossman, T.G.A. (2001) 'Causes of the Decline of the Business School Management Science Course', *INFORMS Transactions on Education*, vol. 1, no. 2, pp. 51-61. http://dx.doi.org/10.1287/ited.1.2.51

Hardin, J.R., Holder, A. Beck, J. Ch., Furman, K., Hanna, A., Rader, D. and Rego, C. (2012) 'Recommendations for an Undergraduate Curriculum at the Interface of Operations Research and Computer Science', *INFORMS Transactions on Education* vol. 12, no. 3, pp. 117-123, http://dx.doi.org/10.1287/ited.1110.0080.

Jacobs, L., C. (1991) *Test Reliability,* [online], Available: http://www.indiana.edu/~best/test_reliability.shtml [3 Aug 2013].

Jarkovská, M., Kučera, P., Vostrá Vydrová, H. and Varvažovská P. (2012) 'Analysis of Students´ Results in Distance-studies centres'. *Journal on Efficiency and Responsibility in Education and Science*, vol. 5, no. 2, pp. 78-91. http://dx.doi.org/10.7160/eriesj.2012.050203.

Jordan, E., Lasdon, L., Lenard, M., Moore, J., Powell, S. and Willemain, T. (1997) 'OR/MS and MBAs', *OR/MS Today*, vol. 24, no. 1, pp. 36-41.

Kaspříková, N. (2011) 'Multivariate Analysis of Examination Papers', *Proceedings of the 8th International Conference on Efficiency and Responsibility in Education,* Prague, pp. 120–127.

Kaspříková, N. (2012a) 'Data analysis of students' performance', *Proceedings of the 9th International Conference on Efficiency and Responsibility in Education,* Prague, pp. 213–218.

Kaspříková, N. (2012b) 'Statistical Evaluation of Examination Tests in Mathematics for Economists, *Journal on Efficiency and Responsibility in Education and Science*, vol. 5, no. 4, pp. 203-211. http://dx.doi.org/ 10.7160/eriesj.2012.050403.

Kolari, S., Savander-Ranne, C. and Viskari E.L. (2008) 'Learning needs time and effort: a time-use study of engineering students', *European Journal of Engineering Education* vol. 33, no. 01, pp. 483-498, http://dx.doi.org/10.1080/03043790802564046.

Miller, I. (2012) *Edukometrie*, [online], Available: http://www.miller.wz.cz/ [13 Sep 2013].

Revelle, W. and Zinbarg, R. E. (2009) 'Coefficients Alpha, Beta, Omega, and the glb: Comments on Sijtsma', *Psychometrika*, vol. 74, no. 1, pp. 145–154. http://dx.doi.org/10.1007/s11336-008-9102-z.

Rydval, J. and Brožová, H. (2011) 'Quantification of Framing effect in education Process using ANP', P*roceedings of the 7th International Conference on Efficiency and Responsibility in Education,* Prague, pp. 36-45.

Škaloudová, A. (2012) *Měření reliability*, [online], Available: http://userweb.pedf.cuni.cz/~www_kpsp/skalouda/pokrocili/reliabi.htm [18 May 2012].

Škoda, J., Doulík, P. and Hajerová-Müllerová, L. (2006) *Zásady správné tvorby, použití a hodnocení didaktických testů v přípravě budoucích učitelů*, [online], Available: http://cvicebnice.ujep.cz/cvicebnice/FRVS1973F5d [25 May 2012].

Tversky, A. and Kahneman, D. (1981) 'The framing of decisions and the psychology of choice', *Science*, vol. 211, no. 4481, pp. 453-458. http://dx.doi.org/10.1126/science.7455683.

Wells, C. S. and Wollack, J. A. (2003) 'An Instructor's Guide to Understanding Test Reliability', *Testing & Evaluation Services, University of Wisconsin*, [online], Available: http://testing.wisc.edu/Reliability.pdf [8 Mar 2013].

Appendix

	2000/2001	2001/2002	2002/2003	2003/2004	2004/2005	2005/2006	2006/2007	2007/2008	2008/2009	2009/2010	2010/2011	2011/2012	2012/2013	Summary
Number of Regular students	99	120	117	155	207	204	225	218	214	181	143	105	185	2173
Unrated	8	17	21	31	44	47	75	45	35	29	16	9	28	405
1 - excellent	14	12	19	16	19	9	17	16	7	5	8	5	2	149
2 - very good	28	24	18	33	42	44	20	41	18	22	23	16	22	351
3 - good	37	58	53	59	78	80	80	73	83	62	71	41	59	834
4 - unsufficient	12	9	6	16	24	24	33	43	71	63	25	34	74	434
Number of attempts	140	169	151	197	261	250	249	332	349	313	255	174	317	3157
Average grade	2.29	2.49	2.38	2.40	2.42	2.53	2.54	2.44	2.70	2.64	2.62	2.58	2.69	2.51
Success rate	0.80	0.78	0.77	0.70	0.67	0.65	0.52	0.60	0.50	0.49	0.71	0.59	0.45	0.61
Average number of attempts	1.54	1.64	1.57	1.59	1.60	1.59	1.66	1.92	1.95	2.06	2.01	1.81	2.02	1.79

	2005/2006	2006/2007	2007/2008	2008/2009	2009/2010	2010/2011	2011/2012	2012/2013	Summary
Number of Distance students	57	62	96	90	93	79	65	73	615
Unrated	25	30	37	33	31	34	5	19	214
1 - excellent	4	6	7	4	3	4	4	3	35
2 - very good	8	8	10	5	6	5	7	3	52
3 - good	13	15	26	20	31	17	17	21	160
4 - unsufficient	7	3	16	28	22	19	32	27	154
Number of attempts	52	53	110	117	106	90	105	93	726
Average grade	2.36	2.31	2.44	2.55	2.70	2.50	2.46	2.67	2.51
Success rate	0.44	0.47	0.45	0.32	0.43	0.33	0.43	0.37	0.40
Average number of attempts	1.63	1.66	1.86	2.05	1.71	2.00	1.75	1.72	1.81

Table 3 Exam results from Student information system (source: own calculation)

Points	2009/2010			2010/2011			2011/2012			2012/2013		
	Number	%	Cumulative	Number	%	Cumulative	Number	%	Cumulative	Number	%	Cumulative
0-29	92	0.283	0.283	65	0.245	0.245	47	0.244	0.244	99	0.339	0.339
30-49	84	0.258	0.541	53	0.2	0.445	50	0.259	0.503	67	0.229	0.568
50-59	59	0.182	0.723	41	0.155	0.6	32	0.166	0.669	48	0.164	0.732
60-73	62	0.191	0.914	67	0.253	0.853	39	0.202	0.871	53	0.182	0.914
74-86	25	0.077	0.991	30	0.113	0.966	21	0.109	0.98	21	0.072	0.986
87-100	3	0.009	1	9	0.034	1	4	0.021	1	4	0.014	1
Sum	325			265			193			292		

Table 4 Frequency of the number of the test points (source: own calculation)

	2009/2010					2010/2011					2011/2012					2012/2013				
	P-value		ULI	Cronbach's alpha		P-value		ULI	Cronbach's alpha		P-value		ULI	Cronbach's alpha		P-value		ULI	Cronbach's alpha	
	All	> 50		All	Second half	All	> 50		All	Second half	All	> 50		All	Second half	All	> 50		All	Second half
Questions	0.389	0.506	0.226	0.586	0.509	0.371	0.664	0.194	0.494	0.514	0.394	0.515	0.267	0.461	0.423	0.358	0.445	0.258	0.549	0.639
Examples	0.448	0.631	0.358	0.448	0.530	0.557	0.998	0.477	0.484	0.503	0.500	0.686	0.422	0.369	0.289	0.450	0.615	0.558	0.547	0.633
Practical example	0.466	0.747	0.710			0.510	0.913	0.667			0.494	0.738	0.615			0.407	0.603	0.582		
Sum of points	0.438	0.640	0.556	0.500	0.566	0.482	0.864	0.803	0.577	0.575	0.466	0.656	0.667	0.512	0.565	0.409	0.560	0.534	0.630	0.659

Table 5 The Difficulty index, Discrimination index and Cronbach's alpha (source: own calculation)