

Intelligent analysis tools for computer based assessments

T. Filatov & V. Popov

Wessex Institute of Technology, Southampton, UK

Abstract

In distance learning the function of measuring the students' knowledge is carried out by computer based assessment systems. However the development of these systems and distance learning as a whole is obstructed by the fact that a human analyst is still required to analyse the reliability of tests used and results obtained.

This study reveals that it is possible by analysing the results of assessments to reveal the qualitative adjectives of the tests. This consequently means that it is possible to implement the self-assessment approach into automated testing systems, which would help teachers to evaluate their tests and assessments results without the help of specialists. This will also provide a platform for further automation and improvement of computer based assessment systems using the same principle.

Following the study, a set of tools was developed aimed at conducting a comprehensive analysis of the results of assessments, to discover the foregoing factors and to report the conclusions in an understandable form.

Keywords: distance learning, computer based assessment, intelligent analysis, statistics, self-analysis, assessment results, test quality.

1 Introduction

Distance learning is developing rapidly nowadays. For the educational system to be really distant and computer based it should have an opportunity to assess students' knowledge remotely in a similar way as delivered. Computer based multiple choice tests are being used for this purpose nowadays providing the opportunity to assess students' knowledge automatically. This makes possible to automate most of the educational process. However the consequences of such



automation could be dramatic because of the influence of human factor during the design of the educational materials and tests. That is why the modern distance learning systems should have the tools for the self-control and human control. These tools should be able to analyze the data collected by the system using the principles of data mining and make assumptions about the positive or negative influence and the quality of different parts of educational process such as different types and sets of learning materials, system interface and usability, assignments, assessment tasks etc [1, 2].

It is possible, we believe, having a considerable amount of data, to distinguish the influence of these factors by a thorough analysis of final results [3]. This analysis can be done in a manual way, in computer assisted way and in a fully automated way. By developing the mechanisms for the computer assisted and automated analysis it is possible to solve the bottle neck problem of the distance learning systems and make the process of education and assessment really automated and reliable [4]. The purpose of this work is to make the first steps in this area. Our task was to develop tools for the intelligent analysis of data collected by the assessment system targeting the quality of the test and questions used.

2 Method

A literature on data mining, statistics, knowledge assessment and theory of measurements was reviewed. A ready computer based assessment system OpenTest was selected which is coded in PHP and stores the tests and results of assessments in the MySQL database which makes it convenient to expand the system with our experimental sub-modules.

- It was decided to develop and integrate the additional analysis tools into the interface of statistics of the selected system.
- The quality of the test and its questions was selected as the main targeted data to reveal.
- A set of intelligent analysis tools was developed and implemented to analyse the qualitative adjectives of the test and its elements. There was also made a successful attempt to unite these tools into one automated tool in order to analyze the quality of the tests automatically.

3 Results

3.1 Questions with deviations

Ideally, each question of the test should have a normal distribution of right and wrong answers by different students. If there is a statistical deviation in the correctness of answers, it is very likely that there is some problem with this question. The tool highlights the answers with statistical deviations making simpler any kind of analysis of questions. The example is shown in Fig.1 where “+” is a column for right answers and “-” is a column for wrong answers.



[the analysis of questions](#) > results overview

test:

Microcontrollers

8-bit microcontrollers of Motorola company						
question	type	alternacy	tries	+	-	+/-
584	1	1/4	8	2	6	0
585	1	1/4	13	4	9	0
586	1	1/4	8	2	6	0
587	1	1/4	11	1	10	0
588	2	2/4	11	3	6	2
589	1	1/4	10	7	3	0
TOTAL	--	--	61	19	40	2

Arithmetical and logical operations in i8x51 microcontrollers						
question	type	alternacy	tries	+	-	+/-
499	1	1/4	8	3	5	0
500	1	1/4	7	4	3	0
501	1	1/4	10	4	6	0
502	1	1/4	10	4	6	0
503	1	1/4	7	3	4	0
504	1	1/4	6	4	2	0
505	1	1/4	5	1	4	0
506	1	1/4	6	5	1	0
TOTAL	--	--	59	28	31	0

Figure 1: Analysis of questions.

3.2 Ability-probability diagram

The idea of this tool was taken from the Items Response Theory [5]. This diagram shown at the picture displays the relation between the knowledge level of the students (ability) and the probability of answering the current question. With a quite high level of reliability the chart displays information about how often the question was answered correctly by students with a high level of knowledge on the subject of a test and about how often the question was answered wrong by students with a low level of knowledge on the topic.

An ideal chart subject to normal distribution of marks should be something like a bias from the bottom left corner to the upper right corner. This will mean that no one student with a zero knowledge level had answered this question and that all the students with a high knowledge level had answered it right. But it is also necessary to take into consideration the factor of normalcy of the distribution, the fact that there is a majority of average levels among students, those who get a "3" mark. It means that in most cases an ideal bias will transform into a curve with a maximum at "3" and a little bend and diminution to "4" and "5" [6, 7].

This tool is very useful for the analysis of separate questions and topicality within the test. It allows one to find out visually whether the question was typical for the test or whether even successful students had some problems with understanding and answering it. If the chart is very untypical, then there is a high probability that the text of the question wasn't understood properly. Example of the diagram: Fig.2.

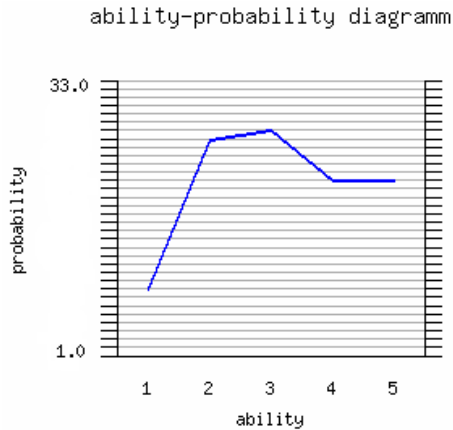


Figure 2: Ability-probability diagram.

3.3 Scale of marks

In the selected computer based assessment system it is possible to set the scale of marks manually. It is known that the distribution of marks for the test should ideally be normal. Therefore, by analyzing the form of the frequency distribution of marks, we can judge about the appropriateness of the scale of marks.

An algorithm of the script of the analysis of the frequency distribution of marks is based on simple statistical formulas:

- 1) the average of distribution for the set of results of the test is calculated

$$\bar{x} = \frac{\sum(x_i)}{n} \quad (1)$$

- 2) the standard deviation is calculated

$$\sigma^2 = \frac{\sum(x_i - \bar{x})^2}{n} \quad (2)$$

- 3) it is known that 34.13% of results when the distribution is normal will lay into the interval between the average of distribution and one standard deviation from both sides of the chart of distribution, two amounts of 13.59% will lay into intervals between one and two standard distributions and two amounts of 2.14% will lay between three and two standard distributions from the both sides of the chart [8].

If the average of distribution is significantly different from the middle of the chart then the chart is shifted and calculations are repeated. If the distribution appears to be normal after that then the conclusions about the difficulty of the test for the students are made. The examples of such conclusions are shown in Fig.3 and Fig.4.

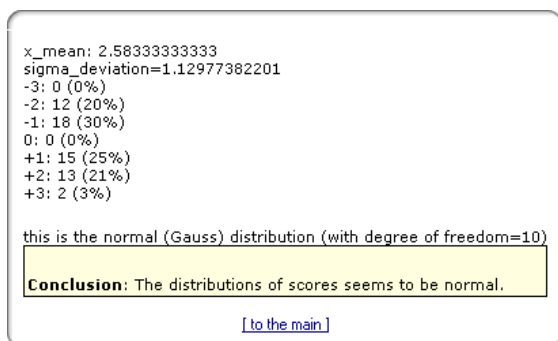


Figure 3: Scale of marks analysis.

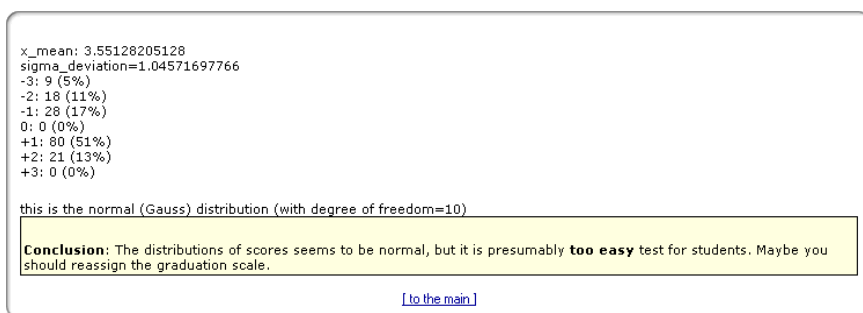


Figure 4: Scale of marks analysis.

3.4 Correlation matrix

The matrix of correlation clearly shows how the answers on particular question had affected the common final result. The source data for it is based on the data of all the students who had passed the current test. The questions that didn't affect the final result and all the suspicious ones are marked with a color. The matrix very clearly demonstrates which of the questions it is necessary to exclude from the test to increase its average discriminative value. The discriminative value shows how good the current test plays his part of classifying the students by their level of knowledge on the particular topic. The summarized and average discriminative values are printed below the correlation matrix and may be used for the control and comparison of the tests.

The coefficients of correlation are calculated by the classical Pearson formulae:

$$r_{xy} = \frac{SP_{xy}}{\sqrt{SS_x * SS_y}} \quad (3)$$

where x – results of students by the current question, y – final results for the test, SP_{xy} – sums of products of deviations from the average values by x and by y; SS_x and SS_y – a sum of squares of deviations by x and by y [8].

This correlation matrix shows how answered questions influence on the total score. If this value is 0 or less, then we can conclude that this question can be excluded from the test (orange colour).

540 0.8	587 0.909	543 0.05	544 0.247	465 0.25	508 2.5	519 0.667	595 1
627 2	531 0	537 2.222	504 1.333	530 0.75	534 0.667	498 1.5	654 0
667 0.667	590 0.5	546 0.8	649 0.75	612 0.8	557 0.357	511 1.5	494 0.667
558 1.333	485 1.429	486 1.429	666 0	584 1.5	583 0.8	548 1.429	560 0.167
488 0	589 2.1	646 0.188	628 1.333	630 0.3	616 1.875	579 1	472 0
459 1.636	525 0.667	606 2.308	471 0.8	601 0.188	665 0.8	599 0	581 0
634 1.2	671 6	521 0.833	610 0.833	559 0	538 2	535 1.5	598 0.833
510 1.636	609 2.5	657 1	618 0.875	622 1.333	638 1.2	561 0.8	484 0
582 0.167	473 1.5	619 0.8	565 0	469 0.429	568 0	629 1.2	576 0.5
522 2.1	509 0.857	573 1.5	507 2.222	577 0.5	621 0	633 1.333	596 0.5
541 1.333	578 1.2	483 0.688	585 2.769	617 0.8	513 2.1	467 0	663 0.667
549 0.875	586 1.5	626 0	503 1.714	644 2	529 1	481 0.667	661 1
487 0.75	614 1.429	463 1.5	464 0.188	550 2	501 2.4	605 0.056	475 0
641 1.2	528 0	554 1.714	623 2.222	591 0	555 1.2	613 0.75	645 0
670 8	562 0.8	600 0.857	480 1.357	493 2	640 0.667	526 1.2	512 0.667
607 2.727	460 3.231	635 1.429	580 0.357	643 2.25	492 1.333	495 0.667	566 0.75
668 1.429	474 0	502 2.4	664 0	552 0.5	514 0.056	631 0.833	497 0
615 0.875	489 0.75	588 2.045	611 2.727	632 0.857	575 1.2	669 0	563 0.125
574 0	625 0.833	518 1.056	647 0.667	527 0.75	533 0.75	477 0	603 0
476 0	637 1.875	536 2.222	490 0.5	515 0.75	542 0	602 0.167	658 0.5
624 2.667	520 2	570 0	547 2.4	659 1.2	553 0.7	539 0.75	556 0.5
468 1.333	594 0.75	470 0.375	642 0.556	524 0	551 1.208	660 0.5	650 0.167
563 1	656 0.667	461 0	652 0.75	466 0.093	462 0	516 0.8	620 0
597 0	639 1.429	564 1.2	604 0.75	571 0	651 0	593 0.5	662 0
636 2.1	499 1.875	500 1.714	523 0.469	517 1.2	592 1.714	506 0.833	648 0
505 0.8	545 1.429	532 1.429	491 0.667	608 1.875	572 0	478 0	655 0

Total test information (discrimination) value: **202.64**
Average test information value: **0.974**

Figure 5: Correlation matrix.

The total test discrimination value is obtained by summarizing the coefficients of correlation of questions. The average discrimination value is obtained by dividing the total value by the number of questions.

3.5 Test quality: automated analysis tool

All the tools listed above will help professionals to produce a high-level analysis of the data collected by the system, giving them powerful and convenient instruments to study the final results fundamentally. Yet the final conclusions must be made by the examiners on their own, and although these tools are applicable and easy to use, they still require a certain training and experience in order to work with them.

The second task of our work was to try and create an automated test analysis tool in order to give to the non-specialists a fully automated tool for the analysis of their tests. This tool should work fully autonomously, using the data of particular assessment and make conclusions by itself. The conclusions should be understandable for humans. The task therefore is to develop an autonomous tool which will replace a human specialist in certain way and make the same conclusions as latter could make after using the above listed tools. An autonomous integral algorithm of the intelligent analysis was developed based on the logical rules. It had embraced all the above described tools. The obtained integral script considers all the necessary criteria such as a number of collected

results, a form of distribution of the marks amongst students, the discrimination ability of the test and values of correlation for every question. Next, the script 'makes' conclusions, understandable for the non-specialist, about the quality of a test, and gives recommendations about the posterior modifications to do to the test.

The example of conclusions made by integral script is presented in Fig.6 and its simplified algorithm is presented in Fig.7.

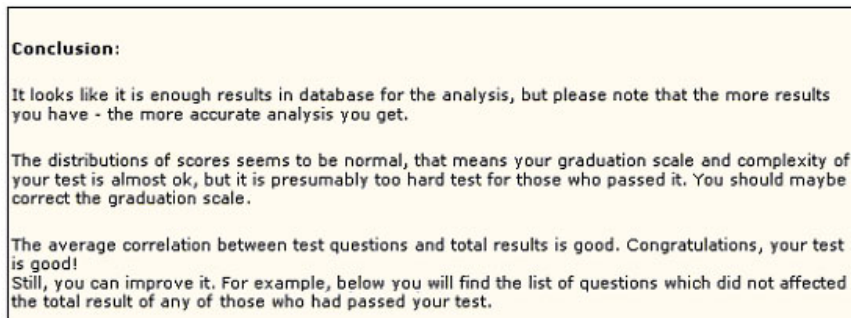


Figure 6: Conclusions made by automated analysis.

3.6 Application so far

The described tools were implemented in the form of additional interfaces for the computer based assessment system. To operate properly, the parameters and degrees of freedom should be configured for each process of integral script. The following parameters were used during our tests:

- >49 (records) in database to proceed with the analysis
- 10% (result points) DOF for the recognition of frequency distribution pattern
- 0.33 (marks) DOF for test difficulty definition
- >0.95 test discrimination value

The results of final university assessments for the courses on 'Informatics', 'Discrete Math', 'Microcontrollers' were taken as the source data. The developed intelligent analysis tools have displayed expected results highly correlated with real expectations. The conclusions given by an integral script of automated test quality analysis were in most cases similar to those given by specialists after a thorough analysis of results. The tools therefore were found very applicable and time-saving by teachers of the appropriate courses.

4 Discussion

A set of methods was developed to help educators to conduct a computer assisted analysis of assessments results. The methods integrated with statistical tools allow discovering non-obvious information about the quality of the test questions and the level of difficulty of the test.



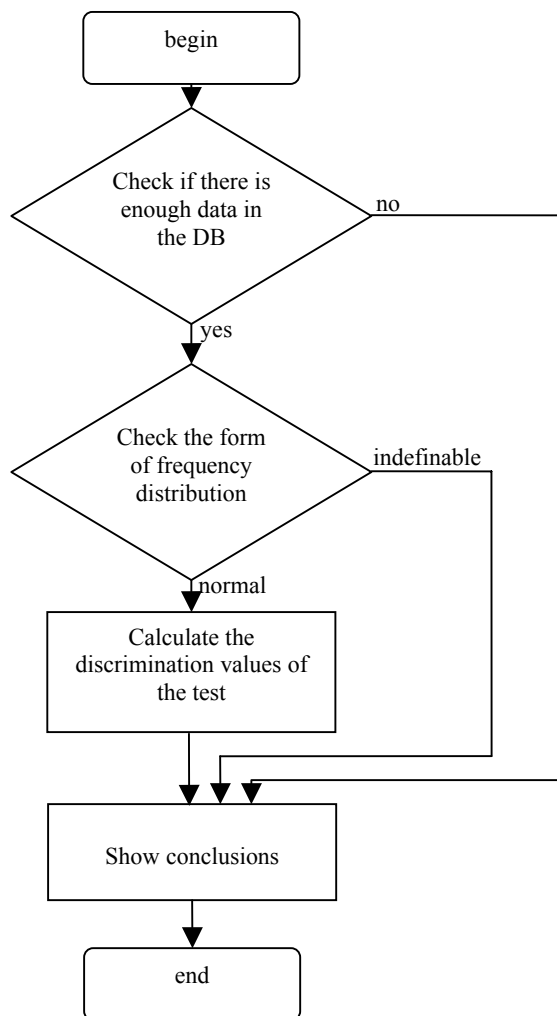


Figure 7: Integral script of automated analysis.

An integral script for the autonomous analysis of the quality of the test was created as a pilot system of a fully automated intelligent analysis tools for distance learning.

The tools were implemented within a real computer-based assessment system and applied for the analysis of results of assessments obtained in real educational processes. This has proved that the developed methodology allows to replace to a considerable degree an experienced human analyst and therefore to economize educators' time.

This represents a base for future investigations and progress in this area. Distance learning systems will become more automated and will save more of

educators' time taking care about the process of learning of each student, allowing educators to dedicate their time to the design of the courses and preparation of new materials. The number of distance learning institutions, the number of courses and the number of students will grow. It follows that there will be a great demand in artificial intelligent systems that will help educators to process and evaluate the data about the progresses of students collected by distance learning systems.

In the current work the achievements of intelligent analysis are applied to evaluate the quality of the test and its contents. Similarly, further research can elaborate methodologies and tools to analyse and evaluate the quality of learning materials, the efficiency of teacher's work, students' knowledge and distinguish these parameters from the collected data using available data analysis methodologies.

References

- [1] Mark Notess (2001) *Usability, User Experience, and Learner Experience*. Indiana University. eLearn. Volume 2001, Issue 8 (August 2001), p.3. <http://www.elearnmag.org/subpage.cfm?section=tutorials&article=2-1>.
- [2] V. Neris, J. Silva, A. Neto, S. Zem-Mascarenhas (2005) *Cognitive strategies to the increase of hyper document quality to distance learning*. Proceedings of the 2005 Latin American conference on Human-computer interaction CLIHC '05, pp. 295-300.
- [3] Usama M. Fayyad, Gregory Piatetsky-Shapiro, Padhraic Smyth (1996) *From data mining to knowledge discovery: an overview, Advances in knowledge discovery and data mining*. American Association for Artificial Intelligence, Menlo Park, CA.
- [4] S.S.J. Lin, E.Z.F. Liu & S.M. Yuan (2001) *Web-based peer assessment: feedback for students with various thinking-styles*. Inst. of Education and Dept of Computer and Information Science, National Chiao Tung University, Journal of Computer Assisted Learning 17, pp.420-432.
- [5] Ronald K. Hambleton, H. Swaminathan, H. Jane Rogers. (1991) *Fundamentals of item response theory*. Sage Publications Inc.
- [6] A. Shkil, S. Chumachenko, S. Naprasnik (2002) *Methodology of evaluation in computer-based knowledge assessment system*. UADL international conference "Virtual education-2002", Yalta, pp.340-345.
- [7] George M. Bodner (1980) *Statistical analysis of multiple choice exams*. Department of Chemistry, Purdue University. Journal of Chemical Education, 57, pp.188-190.
- [8] N. Oleinik (1991) *A textbook on special course "Test as an instrument of measuring knowledge and tasks difficulty in modern educational technology"*. Donetsk State University.

