

The Visualization of Educational Measurement Results in Adaptive eLearning Systems for the Analysis of the Assessment Materials Quality

Yulia Lavdina¹, Oleg Gustun¹ and Evgeniy Antonov^{1,2}

¹ National Research Nuclear University MEPhI (Moscow Engineering Physics Institute), 31 Kashirskoe Shosse, Moscow, 115409, Russian Federation

² Plekhanov Russian University of Economics, Stremyanny lane 36, Moscow, 117997, Russian Federation

Abstract

In the adaptive eLearning system educational measurements are used to control the learning process realized on the feedback approach. Based on the current educational measurements results of the students' achievements the decision is made on the formation of the educational content given to them at the next stage of learning.

To obtain reliable and accurate assessments of the students' educational achievements, it is necessary to use assessment materials that satisfy the requirements for reliability, homogeneity, discriminatory power, validity and other characteristics. In order to evaluate the values of these quality indicators of assessment tools, it is necessary to do sufficient statistical research of the educational measurements results.

At the same time the researcher often encounters the situation when the measurement data are insufficient to achieve the required significance level of the conclusions obtained. In this case assistance in the control decision support can be acquired by analyzing the visual presentation of the educational measurements results.

We propose the visual presentation of the assessment materials characteristics, which helps to evaluate the values of quality indicators of assessment tools during the accumulation of new measurement data. This information combined with a visual analysis of the results of students' achievements makes it possible to identify the trends in the changes of dynamics in students' competencies and to form the educational control action corresponding to their academic performance levels.

Keywords

Educational measurement, adaptive control, visual analysis, quality indicators, assessment materials.

1. Introduction

During the interaction among the participants of the educational process there appear information flows aimed at both sides. A student receives educational materials, tasks and instructions how to do them from a tutor. The student gives the tutor the solutions of the tasks used to put the student a mark. These information flows differ in various characteristics: in the volume, intensity, interactivity and others. The goal-oriented change of these characteristics enables to organize the learning process control.

The change of the educational impact on a student must be made on the basis of reliable information about the student's state. Therefore for the learning process control it is necessary to have appropriate tools to measure the current student's academic performance level.

In accordance with measurement goals educational measurement tools can be divided into two types:

- final measurement for obtaining total characteristics of the student's academic performance after completing the learning process;

GraphiCon 2021: 31st International Conference on Computer Graphics and Vision, September 27-30, 2021, Nizhny Novgorod, Russia

EMAIL: julia_lavdina@mail.ru (Y. Lavdina); gustun@gmail.com (O. Gustun); eantonov@kaf65.ru (E. Antonov)

ORCID: 0000-0002-1188-8119 (Y. Lavdina); 0000-0001-7197-0459 (O. Gustun); 0000-0003-1498-9131 (E. Antonov)



© 2021 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

- current measurement for obtaining feedback during the control of the learning process.

Different types of tools must meet various demands in quality indicators and accuracy indicators.

For the final measurement it is necessary to get valid assessment with high reliability and maximal accuracy. The cost of an error is high for a student here as it is impossible to change the final mark afterwards due to the fact that the learning process has already been completed [1].

Accuracy is not a strict demand during current measurement. The cost of an error is not as high but it is essential to reduce the time of measurement and simultaneously to effectively divide students into classes to which the discriminatory power of the test must conduce.

The values of quality indicators of assessment tools are evaluated on the basis of statistics research of educational measurement results [2]. In this case applied calculation methods depend on the type of measurement instruments and the way of their application, measurement conditions, measuring scales [3].

Researchers quite often face difficulties by interpreting results even in the most typical measurement conditions and simplified calculation methods, for example, when initial data are essentially noisy or insufficient to achieve the required level of conclusions reliability [4]. In this case data visualization can be helpful in both choosing methods of data processing and calculation methods and in support of making decisions during control [5].

2. Adaptive control of the learning process

As a process developing in time, learning can be divided into separate stages. At each stage a student acquires a part of an educational material course. The methodology of acquiring this separate part can recur from stage to stage, it is only the educational content that changes. For instance, at another stage of learning a student gets acquainted with theoretical material of the current course theme (obtains knowledge), does practical assignments (forms competences), gets activity experience (develops skills). At the next stage these steps recur for a new theme. Thus, learning process develops according to the spiral, a student's and a tutor's actions recur cyclically at each new learning stage.

The effectiveness of such learning process depends a lot on the fact how rationally it is organized. Profundity and strength of knowledge acquired by a student, their success in fulfilling assignments, the amount of experience obtained are important for each learning stage. And these characteristics are interconnected.

To influence the total learning process is possible by acting upon separate steps of acquiring a theme at each learning stage.

Profundity of learning theoretical material depends on the amount of educational information and time spent on it by a student. Values of these parameters determine the intensity of learning which can differ for different students.

Strength of theoretical knowledge increases when a student applies them in practice by fulfilling educational assignments. One benefits most of all from completing the assignments which correspond to the student's academic performance level [6].

The influence on learning process effectiveness is achieved by successful fulfilment of the practical assignments difficulty level of which corresponds to the student's academic performance level.

Skills development is evaluated by a student's experience obtained and depends on the time they spent on working with educational content and the number of efforts they made.

The intensity of learning process, the difficulty of educational assignments, the time of working with educational content are parameters which can be changed by acting upon every student individually. In order to evaluate the values of action parameters it is necessary to have a student's model and information about its current state [7].

A student's model must involve the student's knowledge, competences and skills characteristics which are latent and not directly measurable. They can be judged by features that are possible to assess by educational measurements methods.

We will consider the student's model which have the student's academic performance level of solving practical problems as a parameter [6]. Student gain this level after their learning theoretical educational material. The evaluation of this parameter is obtained as a result of educational measuring by a test which is held at each stage of the learning process.

It enables to organize adaptive control of the learning process with the use of feedback [8]. If one uses the difficulty of practical problems as the parameter of control action, the control consists in selecting the assignment whose difficulty corresponds to (or exceeds a little) the student's academic performance level of the course theme given (Figure 1).

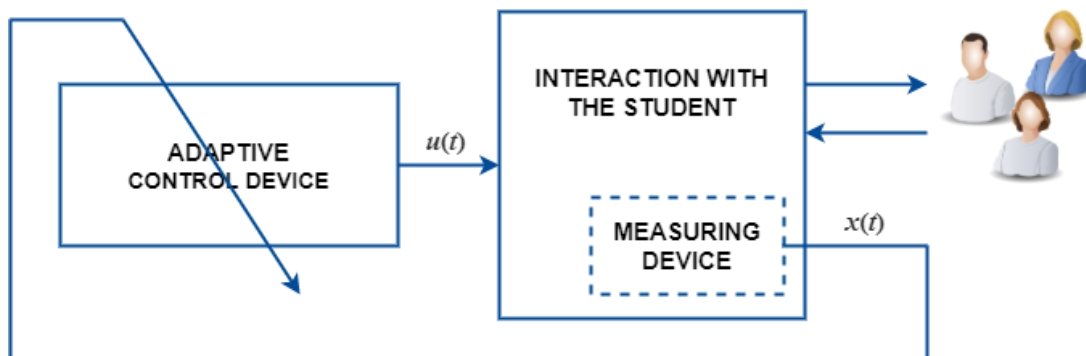


Figure 1: Scheme of adaptive eLearning control system

To realize feedback one uses the measurement device which assesses the student's academic performance level of solving problems during interactive work with the student during computer testing.

3. The quality of educational measurement materials

The main advantages of computer testing as a procedure used for the current measurement of a student's academic performance level are the promptness of conducting an individual test and the rapid processing of its results. Such educational measurements are not difficult to organize for mass use, they can be regularly conducted within minimal educational time [9, 10].

In this article we are analyzing the case when a computer test contains several items with alternative choice of answers presented to a student consecutively. Each item of the test is going to be assessed dichotomously: a student gets score 1 for a correct answer and score 0 for an incorrect one.

Item difficulty and variance are calculated as indicators for items in classical test theory. Item difficulty is defined as the proportion of examinees who answer an item correctly [11]. This is a technical term, which was traditionally formulated somewhat contradictorily: the easier item actually has the higher item difficulty value.

The total score mean and total score variance are calculated for the test which contains several items [12].

Moreover, as a measurement tool the composite test being analyzed must possess necessary quality indices such as reliability, homogeneity, discriminatory power, validity [13]. These indices are calculated on the basis of information about the score distribution of each item and its relationship to other items in the test, they are evaluated by statistical research of the educational measurement results.

The use of visualization methods in analyzing test data by the methods of classical test theory is limited by the fact that researchers mainly build graphs and diagrams based on calculated indicators and characteristics that reflect summarized or complex properties of the objects of research.

A large amount of source and intermediate numerical data are not taken into account in the analysis, since researchers find them difficult to illustrate, perceive and interpret. This leads to the fact that information that may be important and interesting for the researcher eludes the analysis.

Below we give the examples of tables of source data, for which the values of the summarized indicators are equal, but the quality of composite tests varies significantly.

Unfortunately, most of the researcher's tools do not have a well-developed and well-thought-out visualization functionality. At least in the literature there is no description of such tools. Therefore, we propose an approach to data analysis in which visualization methods are applied both when presenting descriptive statistics data and when performing statistical inference.

When doing research of the educational measurement results during the adaptive control of the learning process, which is proposed in [6; 7], visualization methods make it possible to accelerate the decision-making on controlling action in limited time conditions.

4. The visualization of descriptive statistics of a composite test

4.1. Score matrix

In order to calculate item difficulty and their variance a researcher makes person-item score matrix, each its entry involves score of student i on test item j . The example of matrix shown in Figure 2 is presented according to [11]. In the rows below the matrix the values of item difficulty p_j and their variance σ_j^2 are displayed. In the column which is on the right of the matrix the values of students' total scores for the test r_i are indicated.

		Q $_j$					
		Q1	Q2	Q3	Q4	Q5	r_i
S1		0	0	1	1	0	2
S2		0	0	0	1	0	1
S3		1	1	0	0	0	2
S4		1	0	0	1	0	2
S_i	S5	0	1	1	1	1	4
	S6	0	1	0	0	0	1
	S7	1	1	1	1	1	5
	S8	1	1	0	1	0	3
	S9	1	1	1	1	0	4
	S10	0	0	0	1	1	2
	p_j	0.5	0.6	0.4	0.8	0.3	
	σ_j^2	0.25	0.24	0.24	0.16	0.21	

Figure 2: Person-item score matrix

These additional rows and the column help a researcher analyze test results for both evaluating the test quality and assessing students' academic performance level. But the values in the entries of person-item score matrix themselves do not give illustrative presentation of information contained in the matrix.

Conducting the current test is usually aimed at assessing a student's academic performance level of a local theme of an educational course. Therefore the number of concealed factors characterizing a student and influencing the result of doing the test is quite limited. The analysis of the current test results is often aimed at finding the value of only one latent variable parameter of a student's model.

Then in order to display the matrix we can use the approach applied to constructing Guttman Scales [14], which demonstrate a cumulative effect. Firstly, matrix columns have to be sorted by values of item difficulty (for example, in the descending order of p_j); secondly, matrix rows must be sorted by values of students' total scores (for instance, in the descending order of r_i as well). Thus, the matrix will try to look like upper triangular matrix relatively to the secondary diagonal. The sorted matrix presented in Figure 3 enables to interpret test results more illustratively than the one in Figure 2. Visual matrix perception improves if its entries are colored differently in accordance with the values contained in them (Figure 3, left matrix). The coloring on the right matrix of Figure 3 corresponds to the ideal distribution of matrix elements in the Guttman Scale, areas of 1 are indicated in blue tones, and areas of 0 are indicated in red tones.

Usually the quantity of students who have completed a test significantly outnumbers the quantity of test assignments, therefore person-item score matrix is, as a rule, rectangular and long vertically. At the same time it can have a big number of the same rows. In this case after sorting matrix rows there appear

the blocks of the same rows located together. These blocks can be replaced by one row with the indication of the fact how many times it is found in the matrix. It is also necessary to memorize identifiers of students whose answers are the same. It is convenient to illustrate the number of the same rows in each block by the bar chart located vertically on the right and left of matrix.

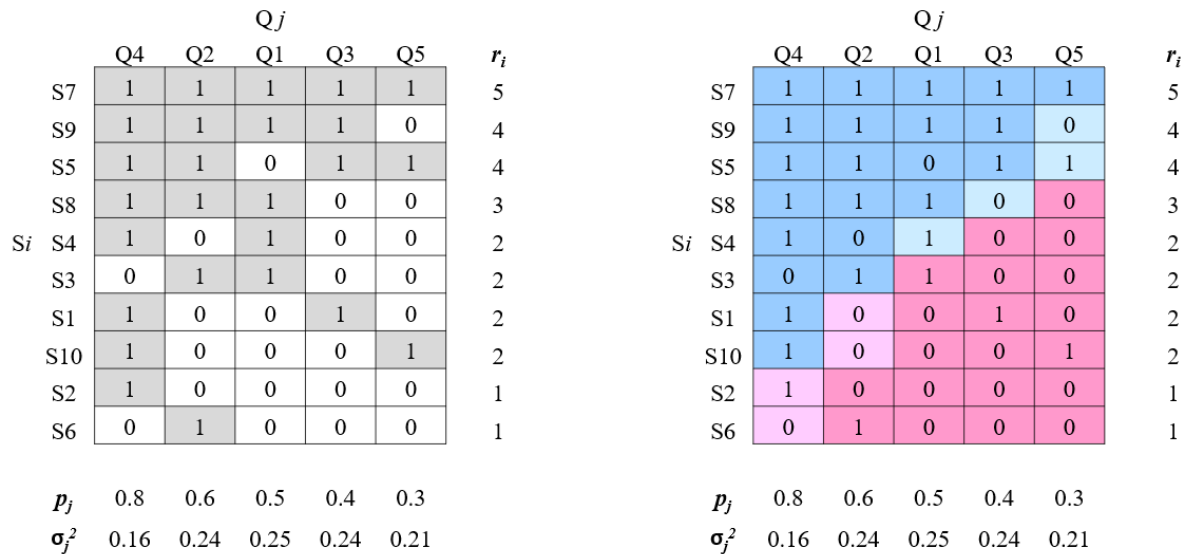


Figure 3: Colored person-item score matrixes

4.2. Variance-covariance matrix

To analyze the total score variance of a test variance-covariance matrix is made. It is the symmetric matrix on the main diagonal of which the values of each item scores variance are located. In the entries there are values of covariance for each pair of test assignments [15].

The sum of values in all of the entries of variance-covariance matrix determines the total score variance of a test so a researcher has to take into consideration the covariances for all different pairs of assignments. If a test consists of n assignments, the number of covariances is equal to $n(n-1)/2$ (for the given example this number is equal to 10). This number grows in quadratic dependence from n .

The situation is getting more complicated due to the fact that the values of covariances cannot be directly interpreted. For this reason it is not easy to conduct the analysis of covariances for a test consisting of quite a big number of assignments [16]. However, this is the task which a researcher has to do.

The reason is that covariances of test assignment pairs can have negative values, which means that assignments in a pair act in an opposed way (if the first assignment is correctly done, the probability of a wrong answer to the second one increases). In this case summed up with other positive covariances total score variance turns to be low.

To visualize the analysis of variance-covariance matrix the heatmap is used when matrix entries are colored in different color palette in accordance with covariance value [17]. It is convenient to use two colors here for positive and negative numbers with their gradations. The heatmap of the variance-covariance matrix for the example analyzed is displayed in Figure 4.

It is useful to demonstrate the examples of person-item matrices (Figures 5 and 6), for which dichotomous values of scores are distributed so that all parameters of difficulty and variance equal the corresponding values of matrices depicted in Figures 2 and 3.

Figures 5 and 6 show two examples of the results of measurement conducted by using two composite tests containing the same number of items. For these tests, both total score means and both total score variances are equal, as well as the difficulty and variance values for each of the corresponding test items. The reliability values of these two tests calculated by using the Cronbach's coefficient are also the same. Only covariances values in the variance-covariance matrixes vary, which is graphically shown in the matrixes heatmaps.

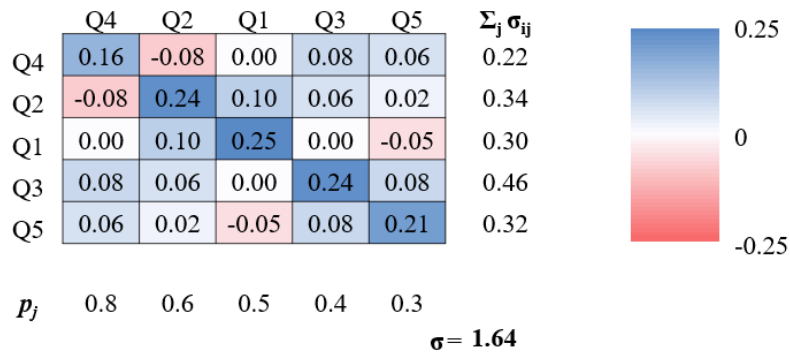


Figure 4: Colored variance-covariance matrixes

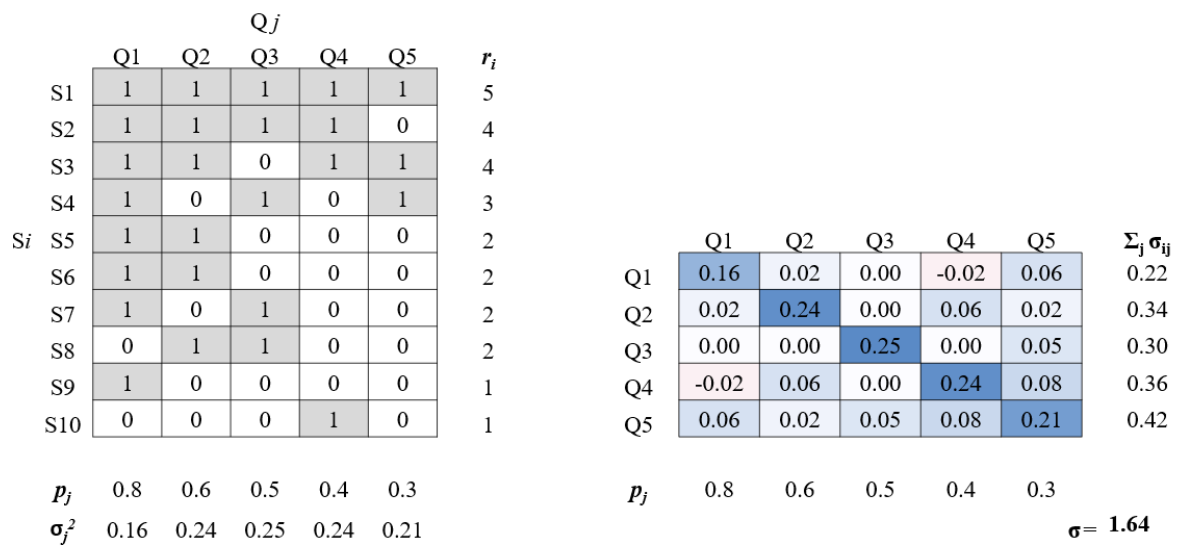


Figure 5: Example 1, person-item score and variance-covariance matrixes

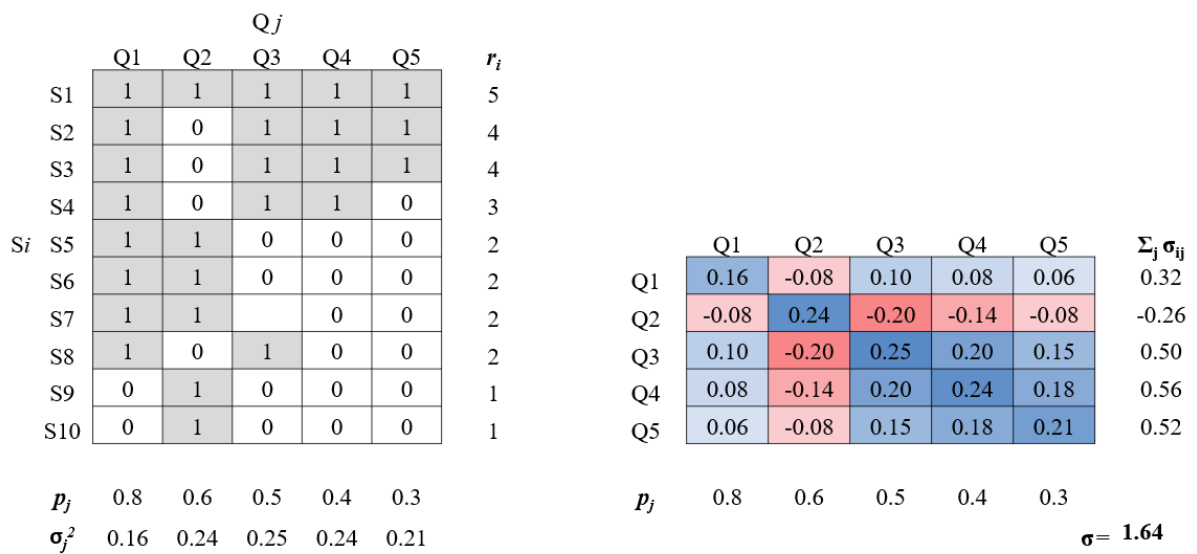


Figure 6: Example 2, person-item score and variance-covariance matrixes

Without the analysis of the variance-covariance matrix, it is impossible to detect a difference of the quality of the tests. For a test with a large number of items, the order of matrix increases significantly, and the advantages of using a heatmap to analyze it become apparent.

Moreover, the heatmap helps when we want to make a new test from separate assignments with familiar parameters of difficulty and variance, which have never appeared in the same test. Then in order to receive the forecast of total score variance of a new test we can use the sorted person-item score matrix by which approximate covariances values are calculated.

5. The visualization of results of statistical inference

Visual analysis of indicators obtained by statistical inference is possible with both the help of earlier described ways of data presentation and on the basis of other methods.

Reliability index expresses the degree of relationship between true and observed scores, and it is determined by using the correlation coefficient of them. The reliability can be calculated by means of the variance of the composite scores and the covariances of the composite test items.

Reliability coefficients are calculated by different method depending on conditions and purposes of testing. For example, Cronbach's coefficient alpha enables to estimate the lower bound of the reliability coefficient for a test on the basis of results of the single administration of the test. To calculate it we can use the values of variances of all items and total score variance of a test:

$$\alpha \geq \frac{k}{k-1} \left(1 - \frac{\sum \sigma_j^2}{\sigma^2} \right), \quad (1)$$

where k is the number of items on the test, σ_j^2 is the variance of item j , σ^2 is the total test variance.

For two tests the matrices of which are displayed in Figures 5 and 6 the values of items variances and total score variance are the same in pairs. It means that reliability coefficients for these two texts are equal. However, the visual analysis of variance-covariance matrices makes it possible to make a conclusion that these tests differ and the measurement can give various errors. For this reason in order to calculate reliability coefficients they usually use several methods at the same time adding the results analysis of descriptive statistics to them. For this reason in order to calculate reliability coefficients several methods are simultaneously used and added by the analysis of the descriptive statistics results.

By analyzing the homogeneity of parallel test forms the results visualization consists in drawing diagrams reflecting the inner features of a test assignment and relations among their different variants.

In Figure 7 there is an example of user-researcher web-interface in which data for the analysis of homogeneity of two different variants of a test item are reflected. The values and diagrams enable a researcher to make the conclusion that characteristics of two variants in the same test are similar and their differences are not statistically significance.

6. The technical implementation of researchers tools

Technical implementation of statistical results visualization, scheme displayed in Figure 8, presupposes an analyst's work with the web application having access to a database of adaptive eLearning system, which collects and stores measurement data.

The PostgreSQL is used as a database for storing students' answers and the main information of the education process. An algorithm of the quality of educational measurement materials is developed in Python programming language and is added to the web application. We used the Apache Superset platform as a visualization tool, it allows us to create the dashboard (presented in Figure 7). This approach was implemented and tested.

To sum up, this stack of technologies provides an analyst opportunity to estimate education materials. The Apache Superset visualization platform is a key tool to find and understand what education materials should be modified.

Tests Groups Students

25 22 419

Binomial Samples

Failed options

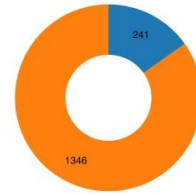
Show All entries

Search 10 records...

id_question	pairs	q
190	f - c	2.16
190	c - h	-2.41
180	c - a	8.47
180	c - b	2.01
180	a - b	-5.57
135	d - a	-3.8
135	d - b	-3.44
135	a - c	6.66
135	b - c	6.1
173	a - b	1.98

Verification result

false true



Homogeneity assessment

Option filter #2

QUESTION_ID

180 x

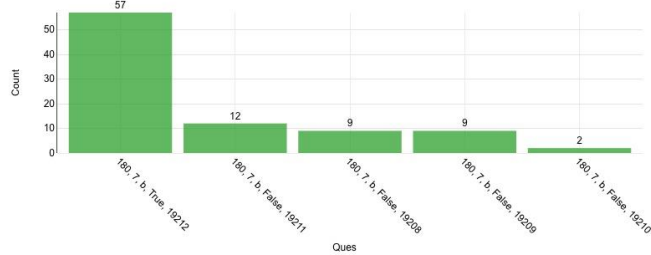
OPTION

b x

APPLY

Answer analysis #1

COUNT(answer)



Option filter #1

QUESTION_ID

180 x

OPTION

c x

APPLY

Answer analysis #2

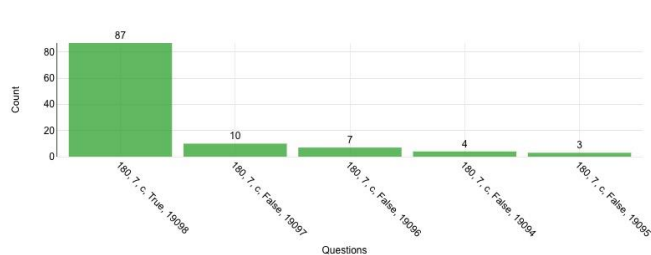


Figure 7: Researcher's interface for the analysis of homogeneity

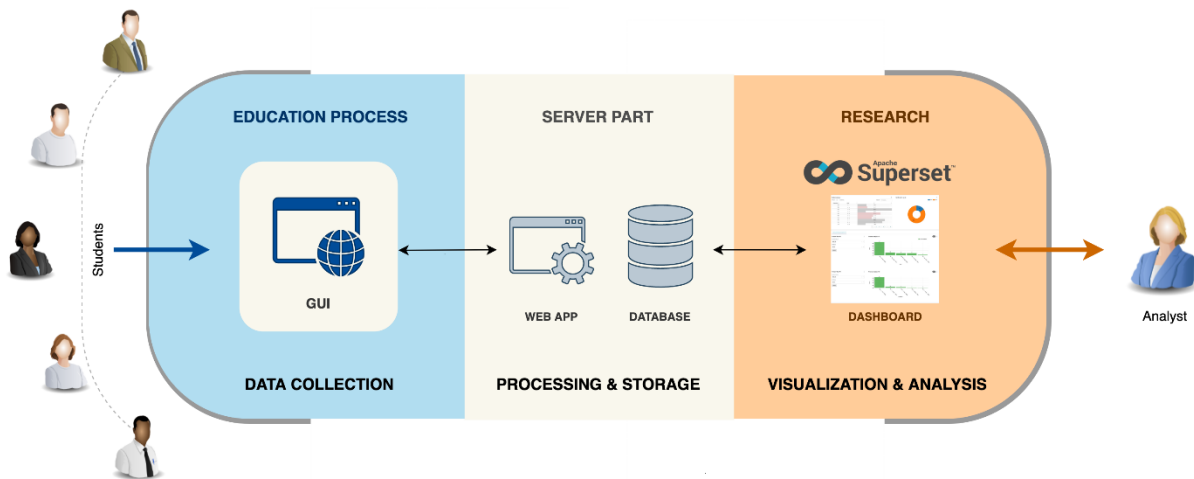


Figure 8: Scheme of technical implementation

7. Conclusions

The choice of data visualization method depends on analysis goals, the researched data character, the content of another stage of analysis. The visualization of calculation results should be used not only for interpreting final results and their evaluation but also between the stages of analysis.

It is useful to order the rows and columns of the person-item score matrixes and check it against the Guttman Scale already at the initial stage of research. If we consider the sorted person-item score matrixes as black-and-white image, then perhaps the use of image processing methods will provide additional information of interest to the researcher.

When calculating descriptive statistics indicators, it is necessary to find the variance-covariance matrix and build a heatmap for it. Without its analysis it is impossible to guarantee the quality and reliability of the composite test.

Visual data presentation helps a researcher put forward hypotheses, check the correctness of models, control conclusions quality, determine further research strategies.

The proposed methods for researching the educational measurements results are suitable for the adaptive control of a learning process, and the developed visual analysis tools make them more accessible to researchers.

8. References

- [1] C. V. Gipps, *Beyond Testing: Towards a Theory of Educational Assessment*, Falmer Press, London, 1994.
- [2] W. Finch, B. French, *Educational and Psychological Measurement*, Routledge, 2019.
- [3] C. Secolsky, D. B. Denison, *Handbook on Measurement, Assessment, and Evaluation in Higher Education*, 2nd. ed., Routledge, 2018.
- [4] D. Cramer, *Advanced quantitative data analysis*, Open University Press, Maidenhead-Philadelphia, 2003.
- [5] R. Mazza, *Introduction to Information Visualization*, Springer-Verlag, London, 2009. doi: 10.1007/978-1-84800-219-7.
- [6] N.M. Leonova, *Sintez algoritmov adaptivnogo strukturno-parametricheskogo upravleniya obrazovatelnoy deyatel'nostyu: Pod red. A.D. Modyaeva, MEPhI, Moskva, 2006.*
- [7] N.M. Leonova, M.V. Markovskiy, *Imitacionnye matematicheskie modeli processov adaptivnogo upravleniya obrazovatelnoy deyatel'nostyu: Pod red. A.D. Modyaeva, MEPhI, Moskva, 2006.*
- [8] S. Sastry, M. Bodson, *Adaptive Control: Stability, Convergence, and Robustness*, Prentice Hall, New Jersey, 1989.

- [9] W. Linden, G. Clas, Computerized Adaptive Testing: Theory and Practice, Springer, 2000. doi:10.1007/0-306-47531-6.
- [10] D. Yan, A. Davier, C. Lewis, Computerized Multistage Testing: Theory and Applications, Chapman and Hall/CRC, 2016. doi:10.1201/b16858.
- [11] Crocker Linda, Algina James, Introduction to Classical and Modern Test Theory, Harcourt Brace Jovanovich, New York, 2006.
- [12] F. M. Lord, M. R. Novick, Statistical Theories of Mental Test Scores, IAP, 2008.
- [13] R. P. McDonald, Test Theory: A Unified Treatment, L. Erlbaum Associates, 1999.
- [14] P. E. Lester, D. Inman, L. K. Bishop, Handbook of Tests and Measurement in Education and the Social Sciences, 3rd. ed., Rowman & Littlefield, 2014.
- [15] Dato N.M. de Gruijter, Leo J. Th. van der. Kamp, Statistical Test Theory for the Behavioral Sciences, Chapman and Hall/CRC, 2008.
- [16] C. R. Reynolds, R. A. Altmann, D. N. Allen, Mastering Modern Psychological Testing, 2nd. ed., Springer, 2021. doi: 10.1007/978-3-030-59455-8.
- [17] P. C. Bruce, A. Bruce, Practical statistics for data scientists: 50 essential concepts, 1st. ed., O'Reilly Media, 2017.