IBSU Scientific Journal, 5(2): 83-92, 2011
ISSN: 1512-3731 print / 2233-3002 online

Basic Procedures of Statistical Quality Control in Education

Irina MILNIKOVA

Abstract

General conceptions of quality management in education are discussed. It is shown that to make decisions of quality control more reliable and objective quantitative evaluation approaches should be used. The approach should be consisting of two stages: 1. elaboration of unitary scaling systems for students grades and 2. clear definition of quality standards. The first stage can be realized by means of equating methodology, particularly by means of equipercentile equating whereas the second one – by means of methodology of contingency tables. The methods allow: 1.comparing of distributions of grades of different subjects in different times and 2. on the base of historical (for the given educational institution) teaching level data specifying quality standards and their confidence levels.

Keywords: statistical quality control, scaling, quality standards, equipercentile equating, contingency tables
JEL Classification Codes: I29

Irina Milnikova *is a doctorate student at Georgian Technical University, Georgia*

1. Introduction

1.1. General conceptions of quality management in education.

Quality management of education investigates whether the process of activity is efficient (whether the goals are achievable). In other words, quality management checks whether relevant systems and structures within organization support the goal of instruction (Hernon, 2002). General purposes of Quality Management in Education are formulated as (Barnes and Wormer, 2005):

· What key outcomes has an educational institution achieved?

· How good is its delivery of education processes?

· How good is its management?

· How good is its leadership?

· What is its capacity for improvement?

According to Harvey and Green (1993) each of these high-level questions can be answered by means of 3 basic actions: 1. definition of what quality is; 2. definition what assessment standards are; 3. comparing the latter with the real outcomes and decide to what extent the standards are met. Thus, one can conclude that quality management is a system that checks whether the produced product or offered service meets the set standards. This approach anticipates three prerequisites: quality is definable, education level index and quality are interrelated and quantitative measurement and assessment of quality is possible. Note that the latter is very important for objectives of the present article.

Also Hernon states that quality assessment should meet the needs of people who benefit from this, as one of the aims of the assessment should be the improvement of activity within the institution under assessment. The advantages of using of quality management in educations can be summarized as follows (Barnes and Wormer, 2005):

- Clarity of Organizational Purpose and Direction
- Higher Student Performance and Lower Dropout Rates
- Roadmap to Achieve the National Education Strategy
- Better Performance Cost Index
- Enhanced Product and Service
- Top Box Customer Satisfaction
- Higher Faculty and Staff Well-Being, Satisfaction, Motivation, and Retention

1.2. Definitin of the Problem.

Quality management has several dimensions. One of them is organizational issue, efficient implementation of which requires quantitative measurement and assessment of quality. The latter implies necessity of usage of Statistical Quality Control (SQC), which makes the whole process of quality management more objective, unbiased and measurable. By its nature SQC can provide quantitative estimation of above mentioned characteristics of educational process: 1. interrelation between education level index and quality; 2. quantitative measurement and assessment of quality. At the same time using of this method in educational processes is in the conceptual stage. Published works consider direct usage of Control Charts (CC) to evaluate development of the learning process in a classroom (Standards and Guidelines for Quality Assurance in the European Higher Education Area(2005); De Feo and Barnard(2005); Oakland(2002)). To understand problems connected with using of CC in the learning processes we represent very short revision of different types of existing CC.

The most widely used CC can be divided into several types (Shewhart, 1931).

CC for Attributes. This type of CC is divided into 2 subtypes: CC for Fraction Nonconforming and CC for Nonconformities. Both of them serve to detect such defects in quality characteristics of output product, which *cannot be represented numerically*. It should be noted, that this kind of CC are constructed on the base of Binomial Distribution. Final decision made on the base of CC for Attributes is “Yes” (Lot is acceptable) or “No” (Lot is unacceptable). Clear that such method cannot be used for quality control in educational processes because of evident reason: during a semester students are evaluated numerically (different types of grades of: exams, assignments, class activities, home works, presentations etc.).

Another widely used type is CC for variables. Firstly note that, unlike CC for attributes, these CC are based on the Normal Distribution and conception of Background Noise. There are several types of control charts, but the common principal idea of their usage is as follows: having the desirable value of a characteristic (it is considered as a mean value $\bar{X}$ of the characteristic) of the measured process acceptable limits of deviations from the mean $\bar{X}$ are established. Assuming that the cause of the deviations is natural background noise distributed as normal variable the limits can be defined as 3 standard deviations (to each side of normal distribution) from $\bar{X}$[1]. They are so called Natural Limits.

No doubts that this type of CC is very efficient in evaluation output of production processes (it has been proved of very long experience of its usage in this field), where and when output objects are (at least they must be!) absolutely identical. Therefore differences among the output objects can have only random character (in majority of cases, as we mentioned above, the randomness follow Normal Distribution) and they can be estimated by means of traditional statistical methods, which finally leads to the technology of CC.

[1] Total 6 standard deviations or 6σ

Apparently, that group of students under consideration represents assembly of individualities. Therefore, technology of CC cannot be sufficiently used to evaluate learning activity of human entities: more humanitarian, more concentrated on a person, approach should be elaborated for this end. Moreover, in the case of educational processes it is not clear how to define standards, and how to compare current states of the process to predefined standards. It is unlike the quality control of production processes, when standards and permitted deviations (2 or 3 sigma) are specified by means of technological and economical requirements.

It is clear, that SQC should evaluate numerical characteristics of education process which are the students' grades. It means, that SQC must evaluate (according to mentioned in 1.1) assessment standards, provide comparing procedure of them to current outcomes (tests, exams grades) and decide to what extent the standards are met.

So, one can conclude, that relying on analysis of tests quantitative results SQC should solve three important problems: 1. comparison of different sets of grades of various disciplines; 2. quantitative definition of quality standards and 3. measuring of to what extent the standards and current results are close (to what extent the standards are met). Solution of each of these problems requires especial methods and approaches. These methods are: Equating and Contingency Tables.

2. Basic Part

2.1. Equating

We use equating as reliable techniques to compare different subjects tests (exams) results to each other and to standards.

Equating method is originated from the problem of comparison of different sets of test scores of different discipline. Point is that different subjects have different difficulties (different test characteristics), so direct comparisons of two (or more) test results are incorrect. Test equating traditionally refers to the statistical process of determining comparable scores on different forms of an exam (Kolen and Brennan, 2004). Equating methods can be used to adjust for differences in difficulty across alternate disciplines, resulting in comparable score scales and more accurate estimates of students ability, which finally defines quality of education.

Equating procedures are used for two types of student groups: equivalent and nonequivalent.

The equivalent groups consist of either a single group of examinees taking several discipline tests or several groups sampled randomly from a single population and considered to be randomly equivalent.

Nonequivalent groups. Nonequivalency of groups means that we sample from two different examinee populations, and the abilities of these groups must then be predefined. Otherwise comparison is nonconsistent.

According to our objectives we shall consider equating of only equivalent groups.

2.1.1. Types of Equating

We are considering now basic types of equating. Equating with the equivalent groups can be categorized as either linear (mean or linear equating), or nonlinear (equipercentile equating).

We assume the following model. There are one group of students and their grades in two different subjects. Grades are considered as random variables X (subject 1) and Y (subject 2) with certain cumulative distribution functions F(X) and G(Y). Correspondingly, x_i and y_i (i=1,2,...,n), where n-number of observations (number of students)), are the sample values of the variables. Equating function is mutual one-to-one mapping of F(X) into G(Y) and vice-versa. It will have different notations for different equating methods.

2.1.2. Identity equating

The identity equating function id(x) simply reproduces the original score value unchanged

$$id_y(x_i) = x_i \,. \tag{1}$$

It means, that scores of both subjects are assumed to be equal. With small samples identity equating, or no equating, has been recommended over other types (Kolen & Brennan, 2004). The identity function can also be combined with any of the functions described below to obtain the synthetic equating function (Kim, von Davier, & Haberman, 2008):

$$s_y(x_i) = (w - 1)g_y(x_i) + w\,id_y(x_i), \tag{2}$$

where $s_y(x_i)$ is a weighted combination of the generic equating function $g_y(x_i)$ with the identity, and w is a value between zero and one.

2.1.3. Linear equating

Linear equating defines a linear relationship between scores from scores X and Y, based on the mean and standard deviation of each. In other words, the standardized scores, or z-scores, are set equal for all score points i:

$$\frac{x_i - \hat{\mu}(X)}{\hat{\sigma}(X)} = \frac{y_i - \hat{\mu}(Y)}{\hat{\sigma}(Y)} \,, \tag{3}$$

where $\hat{\mu}(X)$, $\hat{\mu}(Y)$ expected values and $\hat{\sigma}(X)$, $\hat{\sigma}(Y)$ standard deviations of $F_1(X)$ and $F_2(Y)$ distributions.

When solved for y_i, the linear equating function $l_y(x_i)$ can be rewritten in slope-intercept form as

$$l_y(x_i) = \frac{\hat{\sigma}(Y)}{\hat{\sigma}(X)} x_i - \frac{\hat{\sigma}(Y)}{\hat{\sigma}(X)} \hat{\mu}(X) + \hat{\mu}(X) \cdot \qquad (3)$$

2.1.4. Mean equating

Mean equating is a simplification of linear where the slope, or ratio of standard deviations, is not estimated but is instead assumed to be 1. Deviation scores across scores are thus set equal:

$$x_i - \hat{\mu}(X) = y_i - \hat{\mu}(Y)$$

and the resulting equating function for equating X scores to Y is

$$m_y(x_i) = x_i - \hat{\mu}(X) + \hat{\mu}(Y). \qquad (4)$$

2.1.5. Equipercentile equating

Let $e_Y(x)$ is a symmetric equating function mapping scores of subject X into scores to subject Y. Also let G* cumulative distribution function of $e_Y(x)$.

The function e_y is *equipercentile equating function* (e.q.f.) if

$$G^* = G. \qquad (5)$$

That is, the function $e_Y(x)$ is e.q.f. if cumulative distribution of scores X converted by means of $e_Y(x)$ into scores of Y is equal to cumulative distribution of scores Y.

One can represent e.q.f. as

$$e_Y(x) = G^{-1}(F(x)), \qquad (6)$$

where G^{-1}- is the inverse of G.

By the symmetry property

$$e_X(y) = F\text{-}1(G(y)). \qquad (6')$$

We have considered four types of equating procedures. A question arises: which of them should be used? The choice depends on particular case: clear, that among all feasible methods the simplest is preferable. The most accurate is equipercentile equating due to its nonlinear character. Efficient usage of the latter requires some additional techniques (for example, smoothing), which will be discussed in the subsequence publications.

In conclusion, usage of equating techniques provides conversion of grades of all courses (delivered in certain department or faculty) into unique scale. This provides their correct and consistent comparison to established standards.

2.2. Contingency Tables.

To elaborate standards we use technology of contingency tables (Andersen, 1974).

The following model is considered: n groups of students (each group represents the same year students) during n years (one group each year) took the same m different courses each year. It means that we have nm distributions of corresponding scores.

Among m courses' scores choose one which will be considered as a basic course (it could be chosen, for example, according to professional importance: calculus - for department of mathematics, object-oriented languages - for department of computers etc.). Then for each of n given years we have to equate scores distribution of each of the rest m-1 courses to the distribution of basic course. It results that we would have m equated distribution of each of n years. Based on the latter we can, using well known statistical methods, calculate for each of these distributions their mean or median percentile ranks[2] (assuming that distributions are normal they are equal) f_{ij} (i-number of a year and j-number of a subject).

These statistics can be represented as a contingency table with two independent modes of classification: one mode is time (years) and the other one-subjects.

Table 1. Contingency Matrix

		Subjects (Classification 2)					
		1	...	j	...	m	Row Mean
Years (Classification 1)	1	$\hat{f}_{11}$	...	$\hat{f}_{1j}$	...	$\hat{f}_{1m}$	$\hat{f}_{1.}$
	...	...	...	...	...	...	
	i	$\hat{f}_{i1}$	...	$\hat{f}_{ij}$	...	$\hat{f}_{im}$	$\hat{f}_{i.}$
	...	...	...	...	...	...	
	n	$\hat{f}_{n1}$	...	$\hat{f}_{nj}$	...	$\hat{f}_{nm}$	$\hat{f}_{n.}$
Column Mean		$\hat{f}_{.1}$	...	$\hat{f}_{.j}$	...	$\hat{f}_{.m}$	$\hat{f}_{..}$ - matrix mean

[2] The *percentile rank* of a score is the percentage of scores in its frequency distribution that are the same or lower than it.

The probability $P_i.$ of category i in classification 1 happening and the probability $P_{\cdot j}$ of category j in classification 2 happening are:

$$P_{i.} = \frac{\hat{f}_{i.}}{n\hat{f}_{..}} \tag{7}$$

$$P_{.j} = \frac{\hat{f}_{.j}}{n\hat{f}_{..}} \tag{7'}$$

Since two classifications (Years and Subjects) are independent, the probability of a cell (i,j) is

$$P_{ij} = P_{i.}P_{.j} = \frac{f_{i.}f_{.j}}{mn(f_{..})^2} \ . \tag{8}$$

Using (8) one can calculate Expected Values of each percentile ranks represented in Contingency table.

$$E_{ij} = (mnf_{..})P_{ij}. \tag{9}$$

Each column of Tab.1 is considered as a samples from the same general population. Thus, each values f_{ij} estimated from sample distributions are random and their means E_{ij}, defined in (9), show the expected values of a subject j in year i which hereafter we consider as corresponding standard.

Having estimators of standards and following traditional SQC ideology, we put a question about these standards' Lower and Upper Control limits. They can be calculated as confidence intervals for normally distributed random variables means:

$$LC_{ij} = E_{ij} - Z_{0.997}\frac{S_j}{\sqrt{n}} \tag{10}$$

and

$$UC_{ij} = E_{ij} + Z_{0.997}\frac{S_j}{\sqrt{n}} \ , \tag{10'}$$

where $Z_{0.997}$- 99.7% percentile of standard normal distribution;

$$S_j^2 = \frac{\sum_{i=1}^{n}(\hat{f}_{ij} - f_{.j})^2}{n-1}$$ - variance of rank percentiles of the year j.

If certain mean (median) grade falls out of the corresponding limits, one can conclude the result is not normal. Possible reasons may be due to: bad teaching methodology, bad textbooks, bad class environment etc. All of these require especial researches, which are out of Statistical Methods scope. The main result of elaborated methodology is that it maintains quality of the educational institution's studying process stable, because it detects and then permits to eliminate educational nonconformings.

We have to underline those standards and their Lower and Upper Limits are defined on the base of historical data (statistics) of the institution. Thus, they are not artificial recommendations from outside of institution, but reflect the traditions of the institution and its steady state pedagogical level.

The Lower (10) and Upper (10′) Control limits with $Z_{0.997}$ correspond to 3σ deviations from expected value, so one can refer them as Natural Control Limits. In the case more stable results are desirable one can use more narrow limits, by changing confidence level in standard distribution percentiles[3]. These limits we can refer as Specification Limits, because, unlike Natural Control Limits, they are defined by requirements specified by the quality office or any other external organization.

Conclusion

According to many authors opinion modern challenges in education require establishing of strict and efficient Quality Management system. Among various measures Educational Quality Management Statistical Quality Control should play the important role, because it makes the whole process of quality management more objective, unbiased and measurable. It provides solution of the following decisive problems in the general education quality management: 1. comparison of different sets of grades of various disciplines; 2. quantitative definition of quality standards and 3. measuring of to what extent the standards and current results are close (to what extent the standards are met). Solution of each of these problems requires especial methods and approaches.

We suggest basic conception of implementation of Statistical Quality Control in education which should be based of such statistical procedures as Equating and Contingency Tables methods. Equating, in particular Equipercentile equating, permits to create common scale for evaluating and comparing various grades of, strongly different by their natures, subjects. Then, Contingency Tables methods permit to detect, taking into consideration many years' statistical data of the educational institution, quantitative standards and their natural and, if necessary, specification Upper and Lower limits. The latter allows to provide stability level of teaching, which was formed during the life-time of the educational institution.

References

Andersen, A. H. (1974). Multidimensional contingency tables. Scand. J. Statist. 1, 115–127

Barnes B.J., Wormer J. W. V. (2005). Applying the Performance Excellence Criteria of the Malcolm Baldrige National Quality Program in the Grand Blanc Community Schools, a benchmark in educational improvement efforts. Lansing, Michigan, Quality Facilitators (www.grandblancschools.org/qip).

[3] One σ corresponds to confidence level of 68% and 2 σ – to 95%.

De Feo, Joseph A.; Barnard, William (2005). JURAN Institute's Six Sigma Breakthrough and Beyond - Quality Performance Breakthrough Methods. New York, NY: McGraw-Hill Professional. ISBN 0071422277.
http://books.google.com/books?id=0VHaTb6LJ4QC&printsec=frontcover&dq=%22six+sigma%22&source=gbs_summary_r&cad=0.

Harvey, L., Green, D. (1993). Defining Quality. Assessment and Evaluation in Higher Education, Vol 18.

Hernon, P. (2002). Quality: new directions in the research. Journal of Academic Liberianship, Vol 28, Issue 4.

Kolen M.J, Brenan R.L. (2004). Test Equating, Scaling and Linking. methods and Practice. Springer,

Oakland, J (2002) Statistical Process Control ISBN 0-7506-5766-9.

Shewhart, W A (1931) Economic Control of Quality of Manufactured Product ISBN 0-87389-076-0.

Standards and Guidelines for Quality Assurance in the European Higher Education Area. (2005). European Association for Quality Assurance in Higher Education. Helsinki, Finland