

Quality Self-Assessment in High Schools¹

L'Autovalutazione della Qualità nella Scuola Secondaria

Giancarlo Ragozini

Dipartimento di Scienze Statistiche, Via L. Rodinò, 80138 NA, giancarlo.ragozini@unina.it

Riassunto: La valutazione della qualità negli istituti di scuola secondaria rappresenta sempre più un momento importante per la *scuola dell'autonomia*. Fra i numerosi approcci possibili, facciamo riferimento a quello della valutazione della soddisfazione delle varie componenti scolastiche. In questo lavoro proponiamo un questionario formulato sulla base di alcuni valori condivisi da insegnanti, studenti e genitori, nonché una serie di indicatori di qualità elaborati a partire dal questionario stesso. A scopo esemplificativo si riportano alcuni risultati di una applicazione sul campo degli strumenti proposti.

Keywords: Stakeholder satisfaction, Likert item, education quality evaluation.

1. Introduction

Due to the new legislature scenario that introduces economic and management autonomy, the quality evaluation in the Italian education system, starting from the elementary school up to the universities, is an emerging field. Much attention has been paid to the subject both by researchers and by Public Administrators. Indeed, the Italian Bureau for Education is still running national projects to test the introduction of quality criteria in schools and is setting up a national bureau devoted to the education system evaluation, named I.N.Val.S.I.

Several approaches have been proposed in literature for quality education assessment, based on different way of viewing education and school. Among the others, we briefly recall the *efficiency*, defined as the management of economic and human resources so that resource wastes are avoided; the *effectiveness*, defined as the highest level of the outcome compared with the available resources; the *quality certification* defined as the compliance to the standard of the ISO 9004 and Vision 2000 system. Some experiences of application of the former and other approaches are described in Allulli (2000), Grando (1999) and Negro (1995).

Usually, quality assessment principles for generic services are applied to education. However, with respect to the other services, it has a considerable social and political worth, so that what is common practice for a quality assessment in other fields can be not adequate. Indeed, in the new context of autonomy laws, the efficiency is a crucial factor, however the school system can be not considered just like a profit or business organization. On the other hand, considering the effectiveness as the main quality factor, only final quantity of the knowledge or skills acquired by the students have to be checked (see the OECD's P.I.S.A. project). However, in such a case, misleading evaluations will be obtained if local differences in social and economic contexts are not taken into account. More adequate assessment tools should compare the final skills for each student with respect to the initial ones. Furthermore, the effectiveness could be also interpreted in terms of scholastic, academic or career success or, for instance, in terms of civil and human values acquired by students.

¹ Work supported by M.I.U.R. funds.

Moreover, all the former approaches do not consider school as a community of people – students, teachers, parents and school staff members- integrated in a particular social context, involving inputs, processes and outputs that are human persons.

In this paper we propose a set of quality measures concerning the perceived, expected and acted quality in: human relationships, the sharing of same objectives and methods, teaching methods and other issues connected to the school life. Our proposal falls in a customer satisfaction framework and it is based on a questionnaire presented in Section 2. The proposed statistical measures, that are easy to interpret for people involved in school, easy to collect and to compute and can be used for the quality self-assessment, are shown in Section 3 along with a real application.

2. A proposal of a questionnaire for the quality self-assessment

Our proposal arises from a co-operation with some high school teachers, members of the AICQ-School Section (Italian Association for Quality Culture), that are involved in the A.I.R. (Autovalutazione d'Istituto in Rete) Project (2000). The A.I.R. is an high school network which adopted an identical quality assessment system based on a set of quality indicators selected from the OECD indicators. The quality factors are related to the context, to processes and to outputs in terms of human, economic and structural factors. The main advantage of A.I.R. quality system consists in the directly coming from the school life, as it is not imposed by other external institutions. Hence, it is well accepted by all the school components and it based on information easy to collect and to elaborate by teachers. However, as all the indicators are strictly metric (such as number of hours devoted to refresher courses, number of personal computers, number of meeting among teachers, ...), they do not take into account important issues related to quality in terms of satisfaction, such as personal and psychological aspects.

Objective of our work is to propose a new set of indicators to deeply analyze the former issues, considering the school as a complex system where people play different roles. First, we note that the term “customer” has a particular meaning in the case of scholastic service. Indeed, the main customers are the students and their parents, but also the entire society. Moreover, the labor market can be considered as final customer of the education system. However, the teachers, that provide the service, are also customer of the school (Negro, 1995). Hence, we believe that the idea of *stakeholder satisfaction* is much more adequate to analyze the quality in the school context. Second, we think that a set of questionnaires for each school components (students, teachers, parents, member of staff) is a more appropriate tool to analyze the expectations about their work and themselves, motivation of their actions, their aspirations for a new way of teaching, the organization and the management, the interpersonal relationships.

The model of school which we will refer to is founded on a set of shared values such as: having common objectives (the school mission) and an open mind towards innovation, taking decisions all together, providing new answers to the context changes, having continuous dialogue among people, planning future actions, searching new methodologies and techniques for teaching, connecting different disciplines to deal with complexity.

On such basis, we prepared a questionnaire for teachers, taking in mind that the other questionnaires for the other stakeholders will contain common parts to compare the answers on

the same matters. The teacher questionnaire has been tested by a pilot survey in 3 Neapolitan high schools² to a random sample of 90 teachers, proportionally selected from the three schools.

The main objective of the questionnaire is to evaluate the quality of a school in term of satisfaction. However, it contains also a set of questions providing a portrait of the school status how it is felt and perceived by teachers.

The proposed questionnaire is organized along the following sections: *personal information* such as gender, age, current teaching position, refresher course attendances; *P.O.F. (School Education Project)* evaluated in terms of clearness, verifiability, feasibility, communication of the main objectives to people involved; *education and teaching aims and aspirations*; *didactics*, with particular attention to teaching planning, methodologies and techniques, evaluated in terms of innovation, perceived effectiveness, conjoint planning across different disciplines and satisfaction for the teaching activities; *evaluation methods* for the initial and the final skills achieved by students, the evaluation criteria and their clearness; *human relationships* among the teacher and his students and their parents, the colleagues, the headmaster and the school staff -questions about the perception of the school atmosphere are in this section; *organization and structure*, such as the work organization, the school management, the building characteristics and the other services provided in the school.

3. The proposal of quality indicators

The questionnaire is composed by questions with close answers, both dichotomous responses, multinomial responses and ordinal-polytomous responses. All these types of questions are mixed together in order to avoid the tedious effect due to the presence of only ordinal-polytomous responses and to collect also information about different issues of school life. However, to construct the quality indicators we will take into account only the ordinal-polytomous questions considered in a score model framework.

Related to the subjects listed in Section 2, we associated a set of so-called Likert items. A Likert item consists of a statement and a number of ordered options, where the subject responds by choosing one of the options (Likert, 1932). The items are scored by assigning the rank number of respondent choice and the global score is the unweighted sum of the item scores. Such a type of score can be considered as a continuous score model (Mellenbergh, 1999).

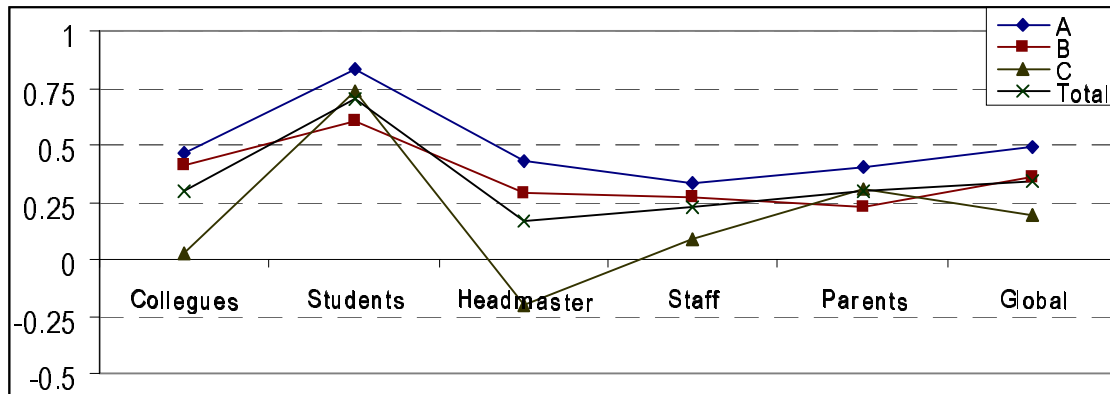
Let X_{ip} the random variable describing the score that the p -th ($p=1, \dots, P$) subject gives to i -th item, $i=1, \dots, n$, with observed values x_{ip} and suppose for simplicity that each Likert items has k ordered options, where 1 represent the worst option and k the best one. Moreover, as in common practice, we assume the distance between two successive options to be constant. The true score is given by the expected value of X_{ip} , $E(X_{ip}) = \tau_{ip}$. It easy to show that the maximum likelihood or least squares estimator for τ_{ip} is x_{ip} . The true global score for the i -th item τ_i can be estimated through the sample mean $\hat{\tau}_i = \sum_{p=1}^P x_{ip} / P$. Under the hypothesis of indifferent evaluation -i.e. people rarely express extreme opinions and the majority of them give medium scores- we should observe a symmetric unimodal distribution centered in $\tau_i^* = E(X_i) = (k+1)/2$. For this reason, we propose as quality indicator the distance between the the expected value of the score hunder the hypothesis of indifferent evaluation and the observed score.

² We thank the teachers L. Iazzetta (AICQ-School Section), A. D'Atri (ITC De Nicola), M. Teti (ITAS V. Emanuele II) e A. Varriale (ITC-Liceo Scientifico "F.S. Nitti") for the co-operation in the questionnaire formulation and in the survey conduction.

Hence, we define the quality indicator for the i -th item $QI^{(i)}$ as $QI^{(i)} = \hat{\tau}_i - \tau_i^*$, with $-[(k-1)/2] \leq QI^{(i)} \leq [(k-1)/2]$. A normalized version $QI_z^{(i)}$ can be obtained dividing $QI^{(i)}$ by $(k-1)/2$, with $-1 \leq QI_z^{(i)} \leq 1$. The quality indicators evaluated in such a way are easy to interpret. Indeed, negative values, close to the minimum, point out some criticisms concerning the quality factor; vice versa, positive values of the indicator show a certain satisfaction degree. Finally, if we consider the sum over all i -th items, we will obtain a total quality indicator defined as $TQI = \sum_{i=1}^n QI^{(i)} / n$. It is worth noting that in such a way all items have same weights in the sum, as usually it is not possible to determine differences among the items.

To exemplify the use of the proposed indicators we refer to the pilot survey described above, considering the set of items related to teacher satisfaction in human relationships. In Figure 1 profiles for the three high schools (here denoted by A, B and C for privacy policy) and total one are plotted. It is possible to note that the global satisfaction degrees are positive, even if far from the maximum 1 (we compute the normalized version $QI_z^{(i)}$). Furthermore, school A achieves always better score than the others, while school C shows a negative value for the relationship among teachers and their headmaster, and lower satisfaction degrees than the other schools. It is also clear that the good relationships with students represent the main reason of satisfaction for teachers. It is possible to interpret the results looking at other items in the questionnaire. For instance, concerning the critical issue in school C related the headmaster, we find that teachers consider him an authoritarian person not open to discussions.

Figure 1: The quality indicators for the human relationships.



References

- A.A.V.V. (2000), *Confrontarsi per migliorare: Progetto A.I.R.*, a cura del Comitato Scientifico Didattico dell'ITCS Primo Levi di Bollate (MI), Il melograno Editore, Milano.
- Allulli, G. (2000), *Le misure della qualità*, SEAM, Roma.
- Grando, T. (1999), *Autovalutazione d'istituto e indicatori di processo*, Trento, IPRASE.
- Likert, R. (1932), "A Technique for the Measurement of Attitudes", *Archives of Psychology*, **140**, 5-55.
- Mellenbergh, G.J. (1999), "Measurements Models", in *Research Methodology in the Social, Behavioural and Life Sciences*, Adèr H.J. & Mellenbergh G.J. Eds., SAGE Publications, London, 168-187.
- Negro, G. (1995), *Qualità Totale a Scuola*, Il Sole 24Ore libri, Milano.

Quasi-experimentation and Educational Research: A Case Study on the Experimentation of Different Statistics Education Strategies¹

*Quasi-sperimentazione e Ricerca in Educazione:
uno Studio su la Sperimentazione di Diverse Strategie Didattiche
per l'Insegnamento della Statistica*

Maria Gabriella Ottaviani
Dipartimento di Statistica, Probabilità e Statistiche Applicate,
P.le A. Moro 5, 00185 Roma,

Silvia Testa
silviatesta@tiscali.it

mariagabriella.ottaviani@uniroma1.it

Riassunto: L'impossibilità di assegnare casualmente gli studenti a strategie didattiche alternative per l'insegnamento della statistica, rende di fatto la sperimentazione effettuata una quasi-sperimentazione. Per effettuare confronti fra i risultati di strategie alternative, si propone allora di utilizzare una procedura di pair-matching che paragona i risultati degli studenti con la distanza di Mahalanobis. I risultati della sperimentazione tuttavia richiedono, per essere interpretati, informazioni sia quantitative sia qualitative.

Keywords: Quasi-experimentation, nonequivalent group designs, pair matching procedure.

1. Introduction

When designing experimental research in education, particular attention must be paid to the procedures involving the creation of control group. In fact, the way by which experimental and control groups are created implies that participants in any experimental research are placed in different treatment conditions and so they are not treated equally. By this ethical problems are originated. Moreover, the fact that the participants are volunteers creates a strong obstacle to the random assignment of participants to groups. Besides this, the experimental mortality - that is the lost of research participants during the course of the experiments - may cause the breakdown of any random assignment situation initially designed. As participants are not randomly assigned to the experimental and control group, experiments in education are not "real" experiments, but they are quasi-experiments

After summarising some quasi-experimentation problems and some possible ways to cope with them, a matching procedure will be applied to a set of data collected in the occasion of an educational research in order to compare different strategies to teach statistics at upper secondary schools.

¹ This paper is financially supported by the grant: MURST 2000, MM13233892_004.

2. Quasi-Experimentation and Statistical Analysis: an Overview

The most important consequence of not random assigning participants to experimental and control group is that "the comparisons from which treatment-caused change is inferred (...) depend on nonequivalent groups that differ from each other in many ways other than the presence of a treatment whose effects are being tested"(Cook, Campbell, 1979, p.6). So the problem is to separate the effects of the treatment from those due to the initial non-comparability between two groups. Distinctions among different types of quasi-experiments have been done. In particular there are nonequivalent group designs and interrupted time-series designs. The latter imply that many measures are taken at many time intervals, before and after treatment. The former are those in which "responses of a treatment group and a comparison group are measured before and after a treatment" (*idem*, p.6). Nonequivalence of control and experimental groups means that the "expected value of at least one characteristic of the groups are not equal even in the absence of a treatment effect." (*idem*, p.148). This implies that "differences resulting from the nonrandom assignment may produce posttest differences between the groups even in the absence of a treatment effect" (*idem*, p. 149). So what researchers need is to separate the effect of the treatment from the effect of the selection differences.

Among the different type of quasi-experimental design, in this paper a nonequivalent control group design is utilised. This is a design in which each group - the control or comparison group and the treatment group - takes a pretest and a posttest that is a test administered prior to or following an experimental treatment. ANCOVA and matching procedures are suitable techniques to analyse the data from a pretest-posttest control-group experiment. ANCOVA allows in fact to compare the posttest mean of the experimental group with the posttest mean of the control group using the pretest as a covariate. The matching procedures allow to select the research participants for the experimental and control group, so to reduce the bias due to non random assignment. This is done matching variables that are supposed to be linked to the treatment variable. Matching variables make explicit some of those "irrelevant causal forces hidden within the *ceteris paribus*" (*idem*, p.6) condition underlined by random assignment.

3. The Data and the Analysis

The National Research Group (NRG) carried out by the University of Padua, Palermo, Perugia and Rome "La Sapienza", and supported by a grant of the Italian Ministry of University and Scientific Research (ex40% 1998) experimented some new strategies to enhance the teaching of statistics at three different school levels: primary, secondary and upper secondary. The NRG followed the international line of thought suggesting that statistics is better taught and learned if real data are proposed in the classroom (Moore, 1997). Thus, in order to teach statistics in the upper secondary schools, three different strategies were proposed: an instructional strategy based on a data oriented approach (DOA); a strategy that accompanies data oriented approach with a computer laboratory as a tool (LAB); an approach where the pedagogical model of Co-operative Learning is added to real data and computer laboratory (CL). The researchers of the four Universities have proposed the mathematics teachers of upper secondary schools to volunteer in the experimental research. The teachers were offered in-service training according to the strategy they had chosen to apply in their classroom. They received

materials, tests and suggestions so that the activities in the classrooms experimenting the same strategy were as similar as possible. From March to May 2000, the participant teachers taught their students the following content units: Statistical information; Statistical survey; A questionnaire for collecting data; Data collection; Frequency distributions; Reading and explanation of tables; Graphic representations; Central measures, Variability measures.

To assess which strategy was more effective the teachers were requested: i) to record the mathematics mark of each of their students at the end of the first four month school period; ii) to submit their students a pretest in order to assess students' skills in reading graphs and tables before the experiment; iii) to submit their students an "intermediate" test to assess their understanding of statistical concepts after being taught the first seven units; iv) to submit their students a "final" test at the end of the experiment.

The analysis of the data at the disposal of the NRG put in evidence that: the experiment made in Rome was significantly different from the experiences of the other places; different strategies had in general different effects, as well as different types of upper secondary schools (Galmacci, Scrucca, 2000). Consequently this papers concentrates on Roman students in order to examine if significant differences exist among strategies. To enhance *ceteris paribus* conditions only a subset of data are analysed. They refer to students of the Technical Institute for Commerce, in class 3rd and 4th. Two data-sets are examined. The first one, referring to the intermediate test, has 276 observations and 5 variables (strategy; mathematics score; pretest score; pretest score/mathematics score; intermediate test score); the second, concerning the final test has 216 observations and 5 variables (strategy; mathematics score; intermediate test score; intermediate test score/mathematics score; final test score). As mathematics scores depends on teacher's way of scoring, the influence of this factor is eliminated considering the minimum and maximum scores utilised by each teacher and introducing the ratios: pretest score/mathematics score and intermediate score/mathematics score. Pretest, intermediate and final test are scored applying to each correct reply a coefficient decreasing according to the item facility. Table 1 shows: the mean scores according to the test and the strategy, and the number of students in each group. It also enable to see that experimental mortality exists and is different for each strategy (DOA -24%; LAB -19%; CL -17%). As the teachers are volunteers and each class must be treated as an "intact group" the one performed is a quasi-experiment.

Table 1: *Mean scores by test and strategy; number of students in each group*

Test	Strategy					
	DOA		LAB		CL	
Intermediate test	40.25	<u>176</u>	36.55	<u>42</u>	46.78	<u>58</u>
Final test	43.16	<u>134</u>	33.70	<u>34</u>	36.63	<u>48</u>

Two questions are suggested by data in Table 1: i) referring to intermediate test, is CL strategy significantly different from LAB? ii) referring to the final test, is LAB strategy significantly different from DOA?

To reply to these questions a matching procedure is utilised. Following Rubin's paper (Rubin, 1973) the pair matching method is chosen. By this, samples are pair matched individual by individual. As three matching variables exist, we propose to pair individuals so that they have comparable pretest scores, using Mahalanobis' distance. The treatment effect is estimated from the posttest difference between pairs. The

analysis reduces to the pair difference t test. The estimate of the treatment effect is the difference of the mean scores of the treatment group and the mean scores of the comparison group.

Referring to the intermediate test, the group of 58 students receiving CL is the treatment group, while the group of 42 students that received LAB is the comparison group. The effect of the treatment is measured on the intermediate test scores after performing the pair matching procedure using the variables: mathematics score; pretest score; pretest score/mathematics score. To remove the effect of co-operative groups one student has been considered for each of the 17 groups. Hence, the samples have 17 pair matched individuals. Table 2 shows that CL strategy is significantly different from LAB.

With reference to the final test, the group of 34 students receiving LAB is the treatment group, and the group of 134 that received DOA is the comparison group. The effect of the treatment is measured on the final test scores after performing the pair matching procedure using the variables: mathematics score; intermediate test score; intermediate test score/mathematics score. The pair matched individuals in each sample are 34. Table 2 shows that LAB strategy is significantly different from DOA.

All the other possible comparisons are not significant.

Table 2: *Strategies comparisons by pair matched samples*

Test	Mean scores		P value of the t test	Number of students
Intermediate test	CL	48.06	0.043	17
	LAB	38.71		
Final test	LAB	33.70	0.007	34
	DOA	44.21		

4. Conclusion

Pair-matching procedure allows to compare the performance of different strategies in this educational quasi-experimental design. Obviously the inferential analysis does not exempt from trying to understand the reason why CL strategy works significantly well in the intermediate test, and DOA in the final test. This may partly be explained looking at the experimentation registers filled by the teachers. They testify that in the final phase teachers had not at their disposals the number of lessons needed to develop strategies other than DOA. By this the needs of quantitative and qualitative information in educational research is made evident.

References

- Cook T. D., Campbell D. T. (1979) *Quasi-experimentation. Design & Analysis Issues for Field Settings*, Houghton Mifflin Company, Boston.
- Galmacci G., Scrucca L. (2000) Strategie didattiche per l'insegnamento della statistica: alcuni risultati di una sperimentazione, Rapporto tecnico, Stesura provvisoria.
- Moore D. S. (1997) New pedagogy and new Content: the Case of Statistics (with discussion), *International Statistical Review*, 65 (2).
- Rubin D.B. (1973) Matching to remove bias in observational studies, *Biometrics*, 29.