



PERCEIVED VERSUS ACTUAL QUALITY IN STUDENT-CENTERED STATISTICS EDUCATION

Patrick Wessa, K.U.Leuven Association, Belgium
Riko van Dijk, K.U.Leuven Association, Belgium

Summary

In this paper the concept Educational Quality is treated from six separate points of view. There are two dimensions that determine each viewpoint: one relates to the measure of quality, the second corresponds to the co-factors that are used to explain and control the outcome

If we are serious about Educational Quality Control then we shouldn't rely on absolute (c.q. uncalibrated) measures (such as Traditional Surveys or Exams) or poorly performing models (such as Satisfaction Models and Test Scores Models). Instead, it is advised to adopt Diagnostic and Engineering Models by introducing technological innovations that allow us to track (and objectively measure) actual learning behavior and appropriately weighted learning outcomes. This allows us to assess Perceived and Objective Quality of learning, taking into account the fact that certain sub populations of students behave in fundamentally different ways and exhibit a diverse array of needs.

Introduction

The concept of the PDCA cycle (Deming, 1986) has been widely applied in Educational Quality Control (EQC). The underlying assumption for the successful implementation of such a cycle is that one is able to clearly define the goals to be achieved, and measure the co-factors that affect the outcomes. Even though EQC may be perceived (by some administrators) as a necessary evil (which is solely associated with the process of external Quality Assurance or "Accreditation"), there are several good reasons to take EQC seriously and treat it as a useful tool for improving education, rather than an obligation.

In this paper the concept Educational Quality (EQ) is treated from six separate points of view. There are two dimensions that determine each viewpoint: one relates to the measure of quality, the second corresponds to the co-factors that are used to explain and control the outcome (Table 1).

Table 1: Overview of quality models

	No co-factors	Reported co-factors	Objective co-factors
Perceived Quality	Traditional Survey	Satisfaction Model	Diagnostic Model
Objective Quality	Traditional Exam	Test Scores Model	Engineering Model

In the following sections, each viewpoint is discussed and related to empirical findings, and some consequences from a pedagogical point of view. Most empirical research results that

are presented have been obtained through the application of a series of experimental statistics courses that have been introduced at the Lessius Business School of the K.U.Leuven Association over the last seven years. This paper provides a review of the research that is related to statistics education at the Business School and highlights the advantages of quality models that are based on objective co-factors.

Traditional Survey

Traditional Surveys have been used at the Business School for many years and focus mainly on student satisfaction about the educator and the course materials. It is obvious that such surveys are virtually useless for the purpose of EQC because the measure of quality (which is supposed to be measured by the survey) is absolute, subjective, biased, and invalid. Absoluteness refers to the fact that the scores that are obtained – typically on a 5 or 7-point Likert scale – have no point of reference. In other words there is no benchmark that allows administrators to make fair comparisons, especially if courses about different subjects are compared (e.g., Language vs. Economics, or Marketing vs. Statistics). The property of subjectivity is caused by the fact that student's opinions are used as a sole measure of quality – without taking into account exam results, prior achievements, or any type of effort exhibited by the student during the semester. Biasedness relates to the fact that the surveys yield scores that consistently deviate from reality due to reasons that have nothing to do with the course that is to be evaluated. Finally, traditional surveys are also invalid in the sense that they do not measure the competences and attitudes that are related to the subject of study and which are deemed to be important by the professional or academic community.

The official surveys which were introduced by the Business School's administration were rather useless due to the fact that they focused on traditional, educator-centered courses where students are supposed to assimilate the information that is provided through lectures (*ex-cathedra*) and a course text that is memorized and reproduced during the examination. Due to the fact that mentioned statistics courses had an experimental nature, with a clearly student-centered focus, a specialized – but well-established – questionnaire (Constructivist On-Line Learning Environment Survey; COLLES) was used to obtain information about perceived student satisfaction and learning experience (Moodle 2008). An interesting property of COLLES is that it incorporates a subjective point of reference: for each of the 24 items in the survey students are asked to provide the preferred level and the perceived level of learning experiences. A comparison of preferred versus perceived experiences has led to the surprising conclusion that our statistics courses result in perceived quality scores that are higher than the preferred scores for certain items. For instance, the degree of tutor-student and student-student interaction (through communication, collaboration, and review) was perceived to be higher compared to what (most students) preferred.

A second survey that was used is the Computer System Usability Survey (CSUQ) which has been developed by IBM (Lewis, 1993) and extended by Poelmans et al. (2008). The extensions of the survey are related to the usability of the web-based system (which includes the statistical software and educational repository) as a tool that allows students to (effectively and efficiently) engage in non-rote learning of statistical concepts. Most notably, some of the added items were used to measure if (and to which degree) there is a comparative advantage of the offered learning environment as compared to traditional ways of learning (e.g., with traditional lectures, handbooks, etc...).

The empirical results of both surveys are discussed in Wessa (2008a). The conclusion is "... that student satisfaction can be very high even though the introduction of the Reproducible

Computing technology and associated Constructivist Pedagogy comes at the price of a heavy workload. In addition, students are able to adopt new learning technologies and engage in many types of technology-based learning activities (such as experimentation, communication and collaboration). ". The results of this investigation are very encouraging but do not lead to any recommendation in regard to the actions that have to be taken (in Deming's Quality Cycle) to maintain or improve EQ.

Traditional Examination

The traditional examination is a (very) poor indicator of EQ because it is absolute, biased, and ill-aligned. The properties of biasedness and absoluteness are similar to the traditional survey – the (only) advantage is that an examination is objective, if the educator uses an appropriate type of examination in which the questions are unambiguous and the answers reproducible. The traditional examination also has the potential to be valid, for instance if the questions attempt to measure true understanding rather than rote memorization. However, even a valid statistical/mathematical examination can be shown to be ill-aligned, which refers to the fact that the relationship between the features and stimuli of the learning environment are not suited to allow all types of students to learn/understand the same concepts to the same degree.

In Wessa (2009a,b) it has been demonstrated that certain types of statistical concepts are not equally well understood by different types of students. For instance, pedagogically substantial, and statistically significant differences were found between female and male students, with respect to the statistical concept of the type II error. Female students were better able to correctly answer the multiple choice questions about the type II error than their male colleagues. On the other hand, male students did a much better job in questions that were related to Spectral Analysis, a statistical technique that involves goniometrics. Other remarkable differences between students with varying degree of prior knowledge were also found in mentioned studies.

Satisfaction Model

A typical example of a model that attempts to explain and predict perceived quality (based on reported data) is related to usability research in which technology acceptance (c.q. quality) depends on survey-based constructs such as: information quality, system quality, ease of use, usefulness, and relative advantage (Poelmans et al., 2008). While this type of research deserves a great deal of attention in the academic literature, the endogenous variable is still absolute, subjective, biased, and (potentially) invalid.

There are various (rather unexpected types of) biases that occur in survey-based measurements as is demonstrated in Wessa (2008b). For instance, it was shown that male bachelor students highly over-estimate their activity-based efforts as compared to their female colleagues. In addition, it was shown that female students from the preparatory programme (they already have a professional bachelor degree) tend to report that they are more focused on communication/collaboration with peers rather than on performing many computations (without social interaction). When these results are compared with actual (objectively measured) learning behavior, one observes that the opposite is true. Overall, students seem to have a poor judgment about their own learning activities and their efforts are related to learning experiences. It is therefore obvious that survey-based measurements are (potentially) highly misleading and not well-suited to perform EQC.

Test Scores Model

In an excellent and extensive study of O'Dwyer et al. (2008), six models were discussed that predicted the Statistics subtest scores of the Massachusetts Comprehensive Assessment System based on a set of reported items - the variance explained ranged between 4% and 7%. Such poor predictive performance is typical for Test Scores Models and clearly illustrates their insufficiency in the context of EQC.

In the past five years, several attempts have been made to build high-quality predictive Test Scores Models based on the data that was obtained in mentioned statistics courses – none of these attempts were successful. For instance, in Wessa (2007) many significant correlations were found between learning outcomes (exam scores) and scores from psychological surveys about student's attitudes and learning experiences. However, it was unclear how these correlations could be effectively used to improve EQ. In addition, it was found that the statistical learning environment was not well-enough aligned with the needs from students to be able to interact with the statistical results that were presented in papers or in the course that was prepared by the educator. In other words, the learning environment did not incorporate the features that allowed users to reproduce, and reuse computations from peers (or the educator) in order to be able to engage in experimentation, collaboration, or review activities.

The above observations have led to the hypothesis that good predictions (about EQ) are impossible to achieve without objective measurements of the actual learning behavior of students. This is only possible if the learning environment allows the educator to accurately measure key-aspects of learning activities for the purpose of empirical analysis. In addition, from the 2007 study it was concluded that the learning environment had to be conceived as a comprehensive system which is not limited to the traditional learning environment (Moodle) but includes the actual statistical software, and computational repository.

Diagnostic Model

There has been a growing interest in using objective measurements of learning activities for the purpose of educational research and EQC. An excellent example of applying so-called Data Mining techniques is provided in Romero et al. (2008) which is based on the Moodle platform. It is clear that such methods will be adopted on a wide scale and emerge in future research.

Based on objective measurements about exogenous co-factors, it is possible to examine certain psychological aspects of learning experiences. The fact that these aspects can be related to objective causes, makes them effective diagnostic instruments. For example, if the methodology of Wessa (2009a) is applied to the data that have been collected in mentioned statistics courses from 2007, it can be shown that male bachelor students' learning satisfaction strongly (and positively) depends on the degree in which they are involved in peer review. This is not the case for female students, and male students from the preparatory programme for whom there are other learning activities that affect satisfaction.

More importantly, not all aspects (items) from the satisfaction survey are affected by exogenous (objectively measured) co-factors for all types of students. For instance, there is a strongly significant relationship between female students' perception that “the tutor models good discourse” and various objective exogenous co-factors. This is not the case for male students. Another example: bachelor students' perception about the question “what I learn is

important for my professional practice” is strongly related to objective co-factors about their learning. This is not true for students from the preparatory programme.

The conclusion is straightforward: different co-factors that are controllable by the educator play different roles in different aspects of student's learning satisfaction. It is impossible to define satisfaction in just one dimension – hence, Perceived Quality is multidimensional and depends on multiple factors.

Engineering Model

The Engineering Model comes closest to Deming's Quality Circle because the endogenous and exogenous variables are (by definition) objectively measured. Within the context of this paper, the Engineering Models are based on innovative statistical learning technology such as described in Wessa (2009d) and are almost identical to the Diagnostic Model except for the endogenous variable which attempts to measure student's competences (exam scores).

An example of such an Engineering Model is given in Wessa (2009c) in which several aspects are described that contribute to EQC. The percentage of variance explained in this model is more than 66% which is considerably better than the Test Scores Models that can be found in recent literature. There are several conditions that have to be satisfied in order to obtain a model with high predictive power. Most notably, the endogenous and exogenous variables must be of high quality (preferably based on objective measurements) and the exam scores must be properly weighted (Wessa 2009a). In addition, there is the requirement that the model is computed for a homogeneous group (for instance, female bachelor students). Alternately, one may compute the Engineering Model for all students simultaneously, taking into account the relevant interaction effects that relate to the group in which students belong (male versus female, and bachelor versus preparatory programme) as is explained in Wessa (2009c). The bottom line is that EQ (and the relationship between EQ with its co-factors) is different for various sub populations.

An illustration of the usefulness of the Engineering Model approach is given by the following quote: “Females who report a high number of submitted feedback messages have significantly lower exam scores. On the other hand, male students who exaggerate their efforts are not in danger of having lower exam scores. This implies that the female exaggeration bias is small but harmful - the male exaggeration bias is big and harmless. ” (Wessa 2009c). This result provides valuable information that allows us to make significant improvements in the statistical learning environment. For instance, female students who tend to exaggerate their own performance should receive accurate and timely feedback about their real performance (as compared to other students) based on objective measures. This form of critical (self) assessment and reflection about student's actual learning efforts should be made an integral part of the statistical learning environment.

Conclusion

If we are serious about EQC then we shouldn't rely on absolute (c.q. uncalibrated) measures (such as Traditional Surveys or Exams) or poorly performing models (such as Satisfaction Models and Test Scores Models). Instead, it is advised to adopt Diagnostic and Engineering Models by introducing technological innovations that allow us to track (and objectively measure) actual learning behavior and appropriately weighted learning outcomes. This allows us to assess Perceived and Objective Quality of learning, taking into account the fact

that certain sub populations of students behave in fundamentally different ways and exhibit a diverse array of needs.

Now is the time for educational scientists and technologists to rise up to the challenge of scientifically sound EQC and embrace the many opportunities that come along with novel techniques and technologies. Our experience and research in the last seven years has shown that it is possible to achieve significant improvements. Policy makers must provide educators with sufficient stimuli to engage in an on-going effort to implement Deming's vision of the Quality Cycle in education.

References

[Deming, W. Edwards](#) (1986). *Out of the Crisis*. MIT Center for Advanced Engineering Study. ISBN 0-911379-01-0.

Lewis, J. R.: IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use, Technical Report 54.786, IBM Corporation, 1993

Moodle: A Free, Open Source Course Management System for Online Learning, <http://www.moodle.org>, 2008

O'Dwyer, L. M., Russel, M., Bebell, D., and Seeley, K.: Examining the Relationship between Student's Mathematics Test Scores and Computer Use at Home and at School, *Journal of Technology, Learning, and Assessment* 6(5), 2008

Poelmans, S., Wessa, P., Milis, K., Bloemen, E., and Doom, C.: Usability and Acceptance of E-Learning in Statistics Education, based on the Compendium Platform, *Proceedings of the International Conference of Education, Research and Innovation, International Association of Technology, Education and Development*, 2008

Romero C., Ventura S. García E.: Data mining in course management systems: Moodle case study and tutorial, *Computers & Education* 51, 368–384, 2008

Wessa, P.: Learning Attitudes, Peer Assessment, and Gender in the context of a Social Constructionist Statistics Course, *Applied Statistics*, 2007

Wessa, P.: Assessment of Reproducible Computing as an E-Learning Tool in Statistics Education, *Proceedings of the World Conference on E-Learning in Corporate, Government, Healthcare, & Higher Education*, 2008a

Wessa, P.: Measurement and Control of Statistics Learning Processes based on Constructivist Feedback and Reproducible Computing, *Proceedings of the 3rd International Conference on Virtual Learning*, 2008b

Wessa, P.: Discovering Computer-Assisted Learning Processes based on Objective Exam Score Transformations, *Proceedings of the World Congress on Educational Sciences*, 2009a

Wessa, P.: How to Evaluate Objective Exam Scores in Statistics Education when Student Populations are Heterogeneous, *Proceedings of the Hawaii International Conference on Education*, 2009b

Wessa, P.: Quality Control of Statistical Learning Environments and Prediction of Learning Outcomes through Reproducible Computing, *International Journal of Computers, Communications & Control* 4(2), 2009c

Wessa, P.: Reproducible Computing: a new Technology for Statistics Education and Educational Research, *IAENG Transactions on Engineering Technologies*, American Institute of Physics, Eds: Rieger, Burghard, Amouzegar, Mahyar A., and Ao, Sio-long, 2009d