

## “Evaluation Traps”: A Brief Vademecum to Avoid the Most Common Mistakes in Distance Learning Evaluation

Gaetano Bruno Ronsivalle<sup>1</sup>, Piera Vivolo<sup>2</sup>

- (1) Professor of University of Rome, La Sapienza  
R&D Manager AbiFormazione, Milan-Rome Italy  
R&D Manager LabelFormazione, Rome Italy  
E-mail: [sprsricercaesviluppo@abiformazione.it](mailto:sprsricercaesviluppo@abiformazione.it)
- (2) Researcher, LabelFormazione, Rome Italy  
80, Scorticabovest, Rome, IT – 00156 – ITALY  
E-mail: [pvivolo@abiformazione.it](mailto:pvivolo@abiformazione.it)

### Abstract

*The purpose of this paper is to draw attention on the methodological inaccuracy often involving the macro-design of evaluation systems for e-learning courses. In our experiences customers often prefer delivering products in record time, to the detriment of appropriate instructional and evaluation design. For example, they ask for web based courses with only summative tests, or for evaluation system design without any content expert's aid; they also often deny the time needed to verify the validity and reliability of the designed instruments. This means that customers underestimate the importance of evaluation by regarding it as something apart from the training program; nevertheless, they always ask for final tests capable of measuring the participants' acquisition of skills and competences. It is, then, necessary to underline the practical problems concerning with the uncontrolled setting-up of evaluation systems from both instructional and methodological point of view. In this paper we propose all the steps of a proper evaluation macro-design activity and then offer some case-studies in order to point out all the problems resulting from omissions occurred in this phase. We describe, for example, what can happen when it is not clear: “what” to measure and “how” to do it, how to design the sampling and to store information to easily build items or how to run empirical analysis. Our final purpose is to create a vademecum to avoid the most common evaluating mistakes and to make customers aware of the practical involvements of a rigorous evaluation system.*

**Keywords:** Assessment, evaluation design, methodology, vademecum.

### 1. The complex concept of evaluation

In learning context, evaluation is a very useful, but multifaceted and complex, instrument. Very few people, among teachers, managers and insiders, are conscious of the traps that evaluation can set. In fact, the more the interest on assessment grows up, the more methodological inaccuracy spreads. Planning and building an evaluation system is not so easy as it comes.

Before going on discussing the many facets of evaluation concept and all its concerning pitfalls, it is useful to briefly analyze the phases of the evaluation process.

Our evaluation system design model consists of 5 main phases:

- **Phase 1:** Design of the *tree of objectives* (cf. Mager, 1975), through the matching between learning objectives and item typologies, on the basis of Bloom's Taxonomy (Bloom, 1956) or other cognitive taxonomies;

- **Phase 2:** Analysis of the "Management variables", that answers to questions such as *why*, *how* and *when* evaluate, and also concerns data selection, report forms and technological restrictions;

- **Phase 3:** Storyboard and items building up, regarding the choice and the design of the proper item typology relating to the specific objective;

- **Phase 4:** Scoring models, concerning also weighted systems and rules;

- **Phase 5:** Review and control of the Beta version test.

Focusing on distance learning, evaluation plays an unquestionable role to guarantee quality processes in learning contexts, for its own features and involvements (*Ronsivalle, et al. 2007*). In spite of this, in our experiences, customers are not usually aware of the importance and significance of a well-done evaluation system: they do not care about its "methodological accuracy" and often ask for assessment only because they need certificates of attendance or because of the general trend.

Unfortunately the problem concerns some instructional designers, too: because of the lack of a theoretical design model, sometimes they design incorrect and inappropriate evaluation systems.

The topic is very tough and ticklish. In the following pages we will discuss some case-studies and show the most common evaluation "traps" at macro and micro design level.

## 2. Common evaluation traps: a vademecum

### I. The ill-described objective trap

**Example:** A food industry requested the design of the evaluation system of a distance training course on "*Meat export Legislation*". During the preliminary analysis, the general objective of the course was not properly focused. Furthermore, the customer asked for a brief testing in order to not tire the end users who were unskilled in pc and new technologies using.

**Trouble:** Items amount was not sufficient nor representative of learning contents, so the test could not measure all the significant objectives.

**Mistake:** Unsuccessful identification of general learning objective and sub objectives. The sub objectives do not reproduce the general objective cognitive difficulty.

**Effects:**

- False positive risk: the training on a particular topic and the attainment of certain competences is certified, but the test measures something else.

- Damage risk for external people and Corporate image: in certain working contexts the lack of competences can be dangerous for the collective safety.

- Money risk: loss of money put in the training intervention.

### II. The protean complexity level trap (or Proteus' trap)

**Example:** A transport industry requested the design of the evaluation system of a distance training course on "*Safety and health at work*". The final users of the intervention were directors, managers and workers. During the macro-design phase, the

different levels of responsibility concerning the diverse corporate roles and the different levels of their expected final achievements were not highlighted properly.

**Trouble:** The achievement test is marked on a particular cognitive level that is suitable for only one kind of user. Items result too much superficial or too much in-depth.

**Mistake:** Lack of adequate identification of items capable of measuring a peculiar learning objective achievement for a specific user.

**Effects:**

- False positive risk: possibility to certify the training on a particular topic and the attainment of certain competences at a definite cognitive level, whereas the acquired level of achievement is higher or lower.

- Damage risk for external people and Corporate image: attainment of competences at a level that is lower than the expected one.

- Money risk: loss of money put in the training intervention with the consequent need of entire re-design of the learning path.

### ***III. The missing goal trap (or Filottete's trap)***

**Example:** A marketing society requested the design of the evaluation system of a distance training course on “*Commercial telephone call management*”. During preliminary analysis, the scope of the testing, that should have been formative instead of summative, was not specified.

**Trouble:** The test does not allow to highlight problems or improvements gained so far. Immediate feedbacks are not provided and users cannot repeat the test.

**Mistake:** Lack of the didactic scope declaration and of the general testing structuring.

**Effects:**

- Testing uselessness: Items result not satisfactory to the formative didactic scope that should focus in-depth only upon some contents.

- Money and time risk: loss of money put in the training intervention and waste of time.

### ***IV. The missing time esteem (Cronus' trap)***

**Example:** An IT society requested the design of the evaluation system of a distance training course on “*European Computer Driving Licence*”. During preliminary analysis, any testing time limit were not established by the customer.

**Trouble:** Platform registered very different average response times among students that ranged from 1'45 seconds to 10 seconds. During the test, users consulted didactic material, even if it was not allowed, and, as a consequence, lengthened their response times.

**Mistake:** Unsuccessful estimation of average duration of test administration and little clarity upon the opportunity of documents consultation.

**Effects:**

- False positive risk: possibility to certify the training on a particular topic and the attainment of certain competences that were not acquired and consequent corporate image damage.

- Invalid test risk and uselessness of initial and final test data comparison.

- Loss of money put in the training intervention and waste of time.

- Need of re-design the training path considering temporal and technological restrictions.

### ***V. The missing selection data entry trap (or Hermes' trap)***

**Example:** A Learning Institution requested the design of the evaluation system of a distance training course on “*English Language*”. Customer asked for final scores reports, without highlighting any particular need to store them or to re-use them in future.

**Trouble:** The users' behaviour on the single item could not be detected. Initial and final administrations reports included only total scores and the correction grid. In spite of platform potentiality, the registered data resulted insufficient.

**Mistake:** Lack of evaluation and selection of the data needed to be registered on the platform.

**Effects:**

- Loss of money put in the training intervention and waste of time.
- Registered data that are not suitable for the setting scope.
- Difficult interpretation of results.
- Impossible data storage and reusability.

### ***VI. The defective monitoring trap (or Argo' trap)***

**Example:** A marketing society requested the design of the evaluation system of a distance training course on “*Communication strategies*”. During preliminary meetings with the customers, the detailed characteristics of monitoring system and data needed to be checked were not defined.

**Trouble:** The LMS platform did not allow to monitor all the concerning data. In particular, it could not manage the weighted questions and the definition of the various interactions.

**Mistake:** Lack of analysis of the technological restrictions of the monitoring system. Missing information for the system developer.

**Effects:**

- Loss of money put in the evaluation system design.
- Unsuccessful management of testing administrations and of data analysis.
- Unsuccessful management of monitoring system.

### ***VII. The defective Reporting trap***

**Example:** An insurance society requested the design of the evaluation system of a distance training course on “*Problem Solving techniques*”. During preliminary meetings with the customers, it was not clear that on the basis of data reports and efficacy index, a rewarding system for the *best achievers* would have been put into operation.

**Trouble:** Data reports resulted insufficient to calculate the Efficacy Index and, so, to define *best achievers*. The report form was inadequate and full of contradictions. The society decided to draw five users among the *High Scorers* at final test.

**Mistake:** Lack of clear and detailed reports.

**Effects:**

- Lot of telephone calls to tutor and protests.
- Corporate image damage risk.

### ***VIII. “The one who think before” trap (or Prometheus’ trap)***

**Example:** A telecommunications industry requested the design of the evaluation system of a distance training course on “*IT safety*”. During preliminary meetings with the customers, it was not clear that they needed an evaluation system with weighted scores and penalties for mistakes occurring in some crucial items.

**Trouble:** Customer needed a weighted evaluation system as soon as possible. On the basis of this request, the computer programmer decided to manage the scoring by himself.

**Mistake:** Lack of clear instructions for the computer programmer about the assignment of weighted scorers and the general scoring rules.

**Effects:**

- Loss of money put in the platform design and programming.
- Assignment of the same score to items of different relevance/difficulty.
- Collection of data that represents only the mere sum of raw scores.
- Corporate image damages.

### ***IX. The “Ambiguous response options” trap (or the Sphinx’ trap)***

**Example:** An Italian bank requested the design of the evaluation system of a distance training course, on commercial topics, aiming at selecting training managers for some new agencies. The customer had little time and did not succeeded in selecting an expert who could validate our items from a content point of view. We had to set up the initial and final tests only on the basis of the materials and contents the bank gave us.

**Trouble:** Users contested the formulation of a multiple choice item. Its three wrong answer options seemed to be very ambiguous and looked all correct.

**Mistake:** Lack of content validity with consequent ambiguity of the items.

**Effects:**

- Lack of satisfactory sampling of contents that the test should measure.
- False positive risk: possibility to certify the training on a particular topic and the attainment of certain competences at a definite cognitive level, whereas the acquired level of achievement is higher or lower.
- Loss of money and waste of time for the re-design of some intervention phases.

### ***X. The “Instructions by heart” trap (or Mnemosyne’s trap)***

**Example:** A food industry requested the design of the evaluation system of a distance training course on “*Spanish Language*” aiming at the selection of some people to be transferred in Latin America for one year. In order to measure all the cognitive levels of knowledge achievement of language (speaking, understanding, writing and pronunciation), we proposed different kind of items (i.e. multiple choice, association, listen & find, etc.). Furthermore, the customer asked for limiting the time of response for each item.

**Trouble:** The instructions for answering the questions were provided only at the beginning of the test. Not recalling all the information, users could quickly solve only multiple choice items. Incapable of giving the right answers to the remaining items, in the established time interval, users asked to nullify the exam.

**Mistake:** Lack of instructions to solve the questionnaire in the right way.

**Effects:**

- Lack of validity and reliability of testing.

- Loss of money and waste of time because of questionnaire re-design.
- Corporate image damages.

### ***XI. The missing clues trap (or Daedalus' trap)***

**Example:** A services company requested to revise the macro-design of its evaluation system of a distance training course on “*Door to door selling techniques* ” in order to select people to engage in this activity. This selection system, in fact, resulted inadequate to the scope.

**Trouble:** The most candidates failed whether in the diagnostic test, or in the final test.

**Mistake:** Lack of a well-designed storyboard with guide lines to design and formulate the items (related didactic objective, cognitive complexity level and item typology).

**Effects:**

- False Negative risk: possibility to certify the lack of training on a particular topic and of the attainment of certain competences.
- Inability to discriminate in the class, the students who really acquired expected competences and those who had not.
- Risk of selecting persons who do not fit the role, or of missing others who do it.
- Loss of money put in the evaluation/selection system.

### ***XII. The novel micro-designer trap (or Hebe's trap)***

**Example:** An e-learning society requested to revise the micro-design of their evaluation system of a distance training course on “*Safety at construction site*”. During the first meeting with the customer, we ascertained that both the training course and the evaluation system were well-designed. In spite of this, the items micro-design caused some problems.

**Trouble:** During the validation phase, the subjects of the random representative sample succeeded in answering well to all questions, both in initial and final tests items.

**Mistake:** Lack of clarity in the setting up of the questions, and, above all, of the answer options that do not result fair attractive.

**Effects:** Loss of money and waste of time because of the needed re-formulation of the items.

### ***XIII. The Item-analysis underestimation trap***

**Example:** A debt collection society requested the design of an evaluation system of a distance training course on “*Aggressiveness Management*”. In spite of the interest showed for all the designing proposals, the customer denied to fund item analysis activities because he considered them pointless.

**Trouble:** The initial and final test means do not differ significantly as expected. Both tests result too much difficult.

**Mistake:** Lack of pre-test and post-test item analysis in order to detect any troublesome item and remove or modify it before testing the users.

**Effects:**

- Lack of validity and reliability of both tests.

- Lack of possibility to detect the level of each user’s achievement with certainty.
- False positive risk: possibility to certify the training on a particular topic and the attainment of certain competences that were not acquired.
- Loss of money, more than the investment required for item analysis.

#### ***XIV. The Omitted comparison trap (or Narcissus’ trap)***

**Example:** An insurance society requested the design of the evaluation system of a distance training course on a new insurance policy. During the first meeting with the customer, his main interest came out: he was, in fact, more interested in certifying that the training had taken place, rather than really training his agents. He asked only for a summative test. After some months, the customer asked data to justify money investment on the training course. Data were, obviously, insufficient for that.

**Trouble:** It was not possible to compare scores obtained by the users at the beginning and at the end of the intervention.

**Mistake:** Lack of design of a diagnostic test to detect previous knowledge and to compare with the final test in order to measure the learning Efficacy Index.

**Effects:**

- Uncertainty of the training path quality.
- Risk of difficult quantification of the amount of added learning value of the course.
- Usefulness of data for methodological analysis.
- Risk to annoy skilled users, by not providing them learning individual paths on the basis of their diagnostic results.

#### ***XV. The fake benefit trap***

**Example:** A services company requested to revise the evaluation system design of a distance training course on “*Personal data management Legislation*”. Despite the design of an initial and a final test in order to compare scores, the analysis of their system highlighted an absolute absence of methodology in the design of the whole training path. In fact, calculated on their data, the Efficacy Index resulted surprisingly negative.

**Trouble:** In spite of improving the knowledge system of the users, the course created only confusion and bewilderment.

**Mistake:** Use of a mistaken correction key that causes the unbalancing scores or, if the worst comes to the worst, absence of intervention and evaluation macro-design.

**Effects:**

- Uncertainty of the training path quality.
- Risk of difficult quantification of the learning efficacy and amount of added learning value by the course.
- Risk of unjustified money investment.

### **3. Conclusion**

In our opinion, in a methodological design approach to evaluation, the common sense is not sufficient. On the contrary, it represents the main trap among the above mentioned ones.

At the same time, we and all insiders know that errors can occur also with a solid design model at the basis.

With this brief vademecum about the most common mistakes, we hope to have highlighted the variables that, according to us, contribute to make a good job and to avoid evaluation traps.

We are going to continue our empirical research in order to investigate solutions and prevent the catastrophic consequences of a mistaken evaluation design.

#### **4. Acknowledgments**

A special thanks to dr.ssa Simona Carta for her revision of the present essay.

#### **REFERENCES**

##### ***Books***

- BLOOM, B. S. (1956), *Taxonomy of Educational Objectives: The Cognitive Domain*, New York: David McKay Co Inc.,
- MAGER, R. (1975), *Preparing Instructional Objectives*, Belmont, CA: Lake Publishing Co.
- ROMISZOWSKI, A. J. (1999), *Designing Instructional Systems*, 7th edition, London.

##### ***Conference Proceedings***

- RONSIVALLE, G. B. (2006), "A model for Quality Assessment, Measurement and Evaluation of E-Learning in the Italian Banking Sector", in *Proceedings of The Fifth International Internet Education Conference and Exhibition*, Cairo, Egypt.
- RONSIVALLE, G. B., CARTA, S., VIVOLO, P. (2007), "New quantitative models and tools for Quality Management in Banking Educational Institutions and Departments", in *Proceedings of The Fourth Annual Conference of Learning International Networks Consortium*, Amman, Jordan.