

From training KPIs to learning KPIs: ensuring effectiveness in learning processes through predictive analytics and data-based tutoring actions

Daniela Pellegrini^{1,†}, Mario Santoro^{2,*†} and Sara Zuzzi^{1,†}

¹Research and Development Division, Piazza Copernico s.r.l., Via Francesco Gentile, 135, 00173, Roma, Italy

²Istituto per le Applicazioni del Calcolo "Mauro Picone" - Consiglio Nazionale delle Ricerche, via dei Taurini 19, 00185, Roma, Italy

Abstract

This work presents the analysis model of the study data available in the LMS platforms specifically designed to analyze potential critical issues as a functional indicator for the possible achievement of the training objectives and completion of the course. The illustrated system highlights how the use of statistical indicators and predictability can be an effective tool for the early identification of possible critical issues in the field of training results, as well as design and organizational inconsistencies that can weigh on the effectiveness of the training system made available. Our work explains how adopting a data analysis model applied to training environments provides the tutoring system with adequate information on potential critical issues to favor targeted interventions on the participants to prevent risks of training ineffectiveness. At the same time, it analyzes the global quality of the courses made available through a perspective of data exploration that starts from the learning experience and enhances the data already present in the LMS platforms.

Keywords

Learning KPI, Critical Issues, Course Quality

1. Introduction

Evaluating learning experiences and identifying areas for improvement are crucial steps in ensuring effective digital training programs. To this end, our research has developed a comprehensive model for analyzing tracking data from the perspectives of digital learning operators, synthesizing all aspects of training into a unified tool. This approach enables us to create a Macro Index of Performance (MIP) and seven sub-indexes: Results, Study Pace, Course Structure, Computer Adequacy, Community Participation, Educational Tutoring, and Process Tutoring. Using feedback from our first experiments, we have oriented our subsequent research toward analyzing critical situations in tracking data. Our novel framework, Learnalizer, combines the MIP and sub-indexes to effectively evaluate online courses, identify areas of riskiness, and provide tutors with potential preventive intervention strategies. This study investigates the effectiveness of Learnalizer through two real-world cases from an Italian public administration, highlighting compelling examples of criticalities and proposed interventions. We aim to demonstrate how Learnalizer can help address participation and performance criticalities in online courses, ultimately contributing to more effective digital training programs. This paper is structured as follows: after a brief overview of the traditional use of KPIs in evaluating training effectiveness and criticalities, we present the Learnalizer framework and its components, including MIP and sub-indexes. We then describe the application of Learnalizer in two real-world cases, highlighting compelling examples of criticalities and proposed interventions. Finally, we summarize our findings and discuss implications for future research.

AIxEDU - 2nd International Workshop on Artificial Intelligence Systems in Education, November 25–28, 2024, Bolzano, Italy

*Corresponding author.

† These authors contributed equally.

✉ dpellegrini@pcopernico.it (D. Pellegrini); m.santoro@iac.cnr.it (M. Santoro); szuzzi@pcopernico.it (S. Zuzzi)

🌐 <https://www.semanticcase.it/> (D. Pellegrini); <https://baltig.cnr.it/users/mario-santoro/projects> (M. Santoro);

<https://www.semanticcase.it/> (S. Zuzzi)

🆔 0000-0001-6626-9430 (M. Santoro)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

2. Theoretical Background

The big data available on the corporate LMS platforms, from training tracking data to enrollment processes and evaluation systems, represents excellent information potential for continuously improving distance learning projects. The training data, however, are often treated from the point of view of training with a significant focus on hours provided, people involved, test thresholds reached, performance levels, and dropout reduction. This focus on training, i.e., on the training offer, represents the point of view of the training institute or company, which indeed constitutes a component of the problem of training effectiveness and knowledge transfer. Correctly, training institutes and companies must be able to evaluate and monitor the training system offered to the participants. However, extensive debates [1] have arisen on how to effectively evaluate training and its effectiveness, from the issues of the ROI of training to the most recent concept of training impact and the modern KPIs. The latter is helpful for those who organize courses to design new content or methodologies or to redefine support models for training activities. In that case, enhancing these improvements based on the users' actual behavior is equally important. It is essential to adopt a learning perspective, that is, to use data analysis to re-read the experiences and behaviors carried out to redesign the training, starting from the actual impacts and experiences of the participants.

Our perspective on data-based decision-making changes significantly when we consider that choices are derived from natural phenomena rather than just good intuitions or professional expertise. This shift in perspective means that our options for redesign and tutoring intervention become user-oriented and needs-oriented. As such, these interventions aim to validate the effectiveness of designed content by considering real-world experiences. Ultimately, this approach enriches didactic and management options by harmonizing them with typical study behaviors.

2.1. The use of data science in training

To respond to the need of a course evaluation and free learning process from standardized formalisms, practices, and customers, one must equip oneself with data analysis tools and intervention strategies. As part of the LearnalyzeR experience [2], we developed a criticality detecting system that aims to support the intervention of tutors day by day through predictive systems, analysis of participation behaviors, and redesign of courses. As a first step, we defined **criticality** as the behavior of a non-neglecting part of participants that leads to a course experience that does not satisfy their fruition, participation, times, and performance needs. We oriented the survey tools toward the recognition of:

- Criticalities from already known patterns using only course data. They are calculated during the course, and the course is already critical when observed.
- Predictive patterns of criticality with respect to historical behavior. They are calculated during the course, and the course is not critical when observed.
- Distorted use of times and verification of units at risk. It can be calculated only at the end of each unit, and then, it can be summarized only at the end of the course.

We will describe our model in details in section 3.

Learning to understand the potential criticalities allows one to adopt a predictive perspective on the problems, reducing the negative impacts of ineffectiveness of the courses, better-managing investments in training and at the same time increase the usefulness and perceived effectiveness and the appreciation of training initiatives. The ROI perspective is managed in progress in terms of containing the risks of ineffectiveness.

Adopting a data analysis perspective also allows to overcome the conformism of the support activities sometimes affected by some typical distortions in the reading of the LMS tracking, as explained in Table 1.

Table 1: Potential risks in reading tracking data

Problem	Description	Data	Actions
<i>Too much generalization</i>	Looking at the global tracking data it is possible to generate false results due to a vision of the global scenario conditioned by some more typical situations. Risk: <i>excessive behaviors categorization</i>	Progress / completions	Different interventions by course progress classes
<i>Naïve approach</i>	Not adopting a data analysis model risks making subjective considerations based on experience and not corroborated by data. Risk: <i>Data selection bias</i>	Times and completions/progress	Reminders on specific cases and notes
<i>Over/underestimate</i>	An intuitive approach to data based on an exploratory method that if not corroborated by appropriate analysis risks to misleading interpretations. Risk: <i>Halo effect of detected problems</i>	Times and completions / progress	Wide and widespread reminders on highlighted case studies

2.2. KPIs from training to learning

Traditionally, KPIs are used to measure the effectiveness of training. The focus was on the added value that training brought to the company. In the literature, the focus is on training that produces enabling results for business performance. The training system is oriented towards producing specific results. This highlights a training perspective approach, i.e. the enhancement of the training offer and its ability to produce results and economic value.

The theme of the evaluation was widely approached already at the end of the 50s with the classic but still significant model of Kirkpatrick [[3]; [4]; [5]; and [6]]. It focused on the theme of the evaluation of the effectiveness of training as well as on the binomial individual-environment, in favor of a logic of sequential 4-step process in which the results of training are progressively brought back by the individual to his context, and therefore require adequate evaluation tools. This dynamic view of the assessment concerns the impacts in terms of:

- **Reactions:** the degree to which participants find the training favorable, engaging, and relevant to their jobs (typically the evaluation of satisfaction of the training experience);
- **Learning:** the degree to which participants acquire the intended knowledge, skills, attitude, confidence, and commitment based on their participation in the training (alias the evaluation of the added value of training, on the cognitive-affective-attitudinal level);
- **Behavior:** the degree to which participants apply what they learned during training when they are back on the job (typically the manifestation of competent behaviors);
- **Results:** the degree to which targeted outcomes occur as a result of the training and the support and accountability package.

There was an intensification of research since the 80s in particular on the issue of the effectiveness of training and its results. For example, the Baldwin and Ford model [7] highlights how learning to be realized needs the interaction between the person, the appropriate structuring of training and an organizational context that provides opportunities to implement learning. Subsequently, Noe's model [8] highlighted how the transfer of individual learning (linked to the acceptance of feedback, individual expectations and an internal locus of control) in the work passes through the support of the organization, also through social support (support for motivation) to implement adequate performance. Thus, andragogic theories highlight how the transfer of learning is strongly linked to the perception of the usefulness of learning to meet career expectations and satisfaction, within a work context oriented to the management of feedback, to a climate of support between peers and superiors, and the opportunities for application of learning [9]. In all these studies the theme of measurement and objectification of the results has always been the reason for greater criticism of the evaluation models. The main criticisms of Alliger and Janak's (1989) concern the sequentiality and the different meanings and values of the various levels of evaluation, with levels 3 and 4 being more important but often not measured for the difficulties of detection [[10]; and[11]]. Further criticism is the inextricable connection between the levels, with the consequent difficulty of isolating and measuring the impacts separately. Subsequently, in parallel with the growth of diversified training experiences (also of an experiential type), a reasoning on the issues of measuring actual results was activated in the literature.

An alternative approach to Kirkpatrick [3]; [4]; [5]; and [6] to measure improvements made by the training function is the method of Robert S. Kaplan and David P. Norton [12] typical of evaluating the performance of organizational units using Balance Scorecards. This model considers four different perspectives:

- economic and financial (economic results that effectively describe the relationship between investment in training and the results generated);
- customer-market (results in terms of satisfaction of internal customers alias participants);
- internal processes (effectiveness of key processes for achieving training objectives);
- learning-growth (generated learning outcomes).

The BSC method (also in the specific case of training) is carried out by identifying KPIs (Key Performance Indicators) in the four areas of investigation. This way, a 360° overall evaluation of the analyzed activity is obtained. From the considerations that emerged over time and with the emergence of the knowledge-based economy and a complex and flexible labor market, continuing education has acquired strategic importance for both individuals and organizations, with the need to understand and enhance the value of human capital. Although conceptually, Kirkpatrick's model maintains its value as a complex conceptual framework of intervention in the training evaluation field, it does require corrective measures in the context of meanings and tools.

My son Jim asked me, How much has your model changed since it was introduced in 1959? I replied that the model remains essentially the same. The concepts, principles, and techniques are as applicable today as when I introduced the model.

[10]; and [11]

If the distinction between learning and behavior is the main merit of Kirkpatrick's model [3]; [4]; [5]; [6], some criticisms arose over time. I.e., that evaluative activities focus on the first level, partly on the second. At the same time, they appear very complex to analyze, and with few and sporadic attempts, the evaluations on the third and fourth (3) levels, moreover, with partly superficial and subjective approaches [13]. From the investigation of Kennedy et al. [14], the problem does not seem to be the lack of interest. Even the training professionals (practitioners) show a broader interest in the higher levels with a lower perception of the usefulness of the lower levels.

The most recent theories look at experimental analysis to identify the measurement of training outcomes and the objective-results comparison through direct evidence between objectives-results [15]; [16]; [17]. Therefore, the models focus on summative evaluation (i.e., tangible, measurable results at the end of the evaluation).

Other interesting models are:

- **IPO Systematic approach** (input, process, output) of Bushnell [18] contends that Kirkpatrick's model focuses only on what happens after the training, but not the entire training process.
- **Brinkerhoff's six-stage evaluation model** [19]; [20] advocated circular evaluation by measuring all the instructional design elements. The Six-Stage Evaluation Model starts with a needs assessment and identifies training goals. Stage two evaluates the program design, and stage three evaluates program implementation, similar to Kirkpatrick's Level 1 evaluation. Stage four evaluates the learning and is identical to Kirkpatrick's Level 2. Stage five evaluates behavior and is similar to Kirkpatrick's Level 3 evaluation. Stage six evaluates how much learning transferred to the results, as does Kirkpatrick's Level 4.
- **Swanson's performance improvement evaluation model** (PLS). Swanson [21] argued that evaluation systems should be directed toward performance improvement regarding three evaluation domains performance (P), learning (L), and satisfaction (S).
- **Holton's three-level HRD** evaluation and research model. Holton's HRD [22]; [23] identifies three outcomes of training – learning, individual performance, and organizational results, all similar to Kirkpatrick's Levels 2, 3, and 4 because Holton stressed that reactions should not be considered a primary outcome of training. It identifies several variables known to affect the effectiveness of a training program.

All cited research highlights the need to focus on the theme of training transfer as generally defined in the literature. The transfer to the work context depends on the level of involvement and motivation of the participant, the quality of training (understood as attention to learning in concrete reference to the activities to be carried out), and the organizational context (understood as openness to practical experimentation of learning). The latest contributions concerning the training evaluation model highlight the importance of a multi-actor approach in which the different interests and potentially conflicting needs of the stakeholders involved, internal and external to the organization, are represented [16].

A limitation commonly indicated by the literature is the need to equip oneself with adequate tools to identify, study and highlight the results brought by the training. So the topic of data is relevant in any training evaluation process. Our work, therefore, is part of the theoretical framework between level 1 and level 2 of the Kirkpatrick model. We will show a model to reduce the potential effects of ineffectiveness due to the training set, the material, or the object of the training, to facilitate a fluid and practical learning experience to maximize training opportunities and then impact organizational behaviors and final results.

3. A model to detect criticality

As extensively described in our previous work [2], we developed a model of analysis of the tracking data starting from the experiences of digital learning operators to create a tool capable of synthesizing all aspects of training. We developed a Macro Index of Performance (MIP) and seven sub-indexes: Results (I_R), Study Pace (I_{SP}), Course Structure (I_{CS}), Computer Adequacy (I_{CA}), Community Participation (I_{CP}), Educational Tutoring (I_{ET}), and Process Tutoring (I_{PT}). Each sub-index consists of variables that describe the various aspects of the sub-indexes.

The MIP (and sub-indexes) makes comparable (online, classroom, blended) courses through a unified analysis model. It can discriminate different experiences and indicate the expected outcomes based on similar data, guiding the tutors in the differentiated intervention methods to support and facilitate learning.

Using the feedback received after the first experiments, we oriented the subsequent research to study critical situations in the tracking. Using the values of the MIP (and sub-indexes) and the raw values of the tracking, we could highlight areas of riskiness and provide tutors with potential preventive intervention strategies. The MIP and sub-indexes describe the students' behaviors and indicate any critical issues that characterize the courses. Through the automated analysis of the courses, tutors can



Figure 1: Typical participation trend in an online course

focus on those who are most at risk for different types of critical issues. Following the general definition of criticalities in section 2.1, we defined three classes of criticalities:

- **Participation (C_{PA})**, the access and use of course materials;
- **Performance (C_{PE})**, the effectiveness of the training course according to the training plan;
- **Structure (C_S)**, criticality in using educational resources with excessive use of time and expenditure of organizational costs.

3.1. Criticality Details

It is essential to define what critical course means concerning a group of participants and for the individual user. Generally, the criticality of a course is considered based on the use of the time allocated and the need to conclude all the teaching activities to reach the minimum requirements. However, a course is not always critical, or at least not for everyone. Generally, there is an increasing participation distribution that peaks on 3/4 of the time available (Figure 1).

Table 2: Criteria for participation criticality algorithms (C_{PA})

C_{PA} TYPE	CONDITION	PROBLEM	INTERVENTION
A	Elapsed available time between 15% and 25%, not less than 75% of participants have a use time of less than 25%	Reduced number of users who have started	Reminder to Start
B	Elapsed available time between 25% and 50%, not less than 50% of participants have a use time less than 50%	Reduced number of users with significant progress	Reminder to Study
C	Elapsed available time between 50% and 75%, not less than 25% of participants have a use time of less than 25%	Risk of drop out	Reminder to Recover
D	Elapsed available time over 75%, not less than 25% of the participants have a use time less than 75%	Users with completing difficulty	Reminder to Complete
X	Available time ended, at least 10% of users did not complete the course	Incomplete course	Extension or Recovery

Table 3: Criteria for structural criticality algorithms (C_S)

C_S TYPE	CONDITION	PROBLEM	INTERVENTION
Z	At least 50% of the participants have fruition times greater than twice the time provided in the course design	Inconsistent use of time, educational resources	Review course rules and/or design

Table 4: Criteria for performance criticality algorithms (C_{PE})

C_{PE} TYPE	CONDITION	PROBLEM	INTERVENTION
A	Elapsed available time between 20% and 40%, less than 20% of participants with MIP forecast at the end of the course > threshold value	Inadequate study behaviour	Encourage effective study
B	Elapsed available time between 40% and 75%, less than 50% of participants with MIP forecast at the end of the course > threshold value	Study behavior at risk of ineffectiveness	Make organizational proposals and give advice
C	Elapsed available time between 75% and 80%, less than 50% of participants with MIP forecast at the end of the course > threshold value	Highly risky study behavior	Give organizational proposals, re-registrations / extensions
X	Available time ended, at least 10% of users with end-of-course MIP below the threshold	Incomplete course	Extension or Recovery

We should use different rules when the course (time) progresses. Using this framework, we built a criticality detection system inside LearnalyzeR (our learning analytics). Concerning the various critical issues based on the problems highlighted, we carried out particular interventions that the tutors and the system can implement to prevent the risk of ineffectiveness (Table 2, Table 3, Table 4).

3.2. Critical, predictive and preventive tutoring models

The course critical issues detection algorithms compare the raw data of use and the MIP of all the course users with what the course designers and commitment expect. It is necessary to distinguish two types of expected behaviors: the one during the delivery and the one at the end of the course. The criticality algorithms C_{PA} (Table 2) and C_S (Table 3) intercept the expected behaviors during the course, even at the beginning. In contrast, the algorithms for C_{PE} (Table 4) are based on the prediction of performance at the end of the course (final MIP) and depend on the value of the MIP at various steps of the course (available time and completion). For this last criticality, we used a Bayesian beta regression model [2], using the minimum of the MIP estimation interval as a precaution. Then we chose the threshold value in Table 4 by analyzing the MIP values in the previously closed courses. In the following examples,

the chosen threshold value for the minimum MIP forecasted is 70¹. The model, i.e., its parameters², are adaptable, through the retrospective analysis of the closed courses, to the customer's requests, the course structure, and the performance requests.

Our tracking critical issues models allow for timely tutoring actions, which, thanks to the predictive model of the final performance, we can define preventive tutoring.

Tracking critical issues during the course also allows the ex-post evaluation of tutoring actions' design, delivery, and validity.

3.3. Data based tutoring interventions

One defined criticality criteria, we must combine specific intervention criteria based on observed data. Table 5, Table 6, and Table 7 show specific intervention actions for critical issues, as defined in Table 2, Table 3, and Table 4, respectively.

Table 5: Interventions proposed for participation's criticalities (C_{PA}) as defined in Table 2.

CRITICALITY	INTERVENTIONS
C_{PA} type A	<p>a1) Verify that everyone received login information;</p> <p>a2) verify with the client and/or the managers that there are no work contingencies (deadlines, unforeseen events, workloads, etc.) that affect the people's availability;</p> <p>a3) check the sub-index of IT adequacy to verify if there are technological barriers that influence the correct performance of the study activities;</p> <p>a4) check the sub-indices of results and pace of study and verify on the platform the actual times recorded;</p> <p>a5) assess people's level of engagement and motivation for the course.</p>
C_{PA} type B	<p>b1) Send reminders to those who have not accessed the platform, remotivating the importance of participating in the course;</p> <p>b2) for those who have started, motivate participation to recover any study gaps compared to the expected;</p> <p>b3) remind everyone of times and deadlines;</p> <p>b4) check the sub-index results (Pace of study and IT adequacy), and verify the raw participation times recorded, signaling users to respect the study times provided to obtain certificates and certifications.</p>
C_{PA} type C	<p>c1) Recover participants who have been left behind, for whom it is necessary to ask for an effort of participation;</p> <p>c2) identify and quantify the cases of proper abandonment of the course;</p> <p>c3) check the sub-index results (Pace of study and IT adequacy), and verify the raw participation times recorded.</p> <p>For people who have lagged in the fruition, i.e. who have not connected or who have short times compared to the expected, it is essential to provide precise indications to complete the course, indicating average daily study time, minimum necessary activities, and operational advice. For people with study times that do not comply with expected, proceed to report that it is essential to respect the study times indicated to obtain certificates and certifications.</p>

¹Recall, as described in [2], the MIP assumes values between 0 and 100

²Percentage of times and users for participation, structure and performance; threshold value of the index for predicting performance.

CRITICALITY	INTERVENTIONS
C_{PA} type D	<p>d1) Communicate to participants that they are at risk of non-completion and non-compliance with the requirements;</p> <p>d2) verify the problems encountered;</p> <p>d3) highlight a possible study plan to recover and complete the course.</p> <p>For people who have lagged in the fruition, i.e. who have not connected or who have short times compared to the expected, it is essential to provide precise indications to complete the course, indicating average daily study time, minimum necessary activities, and operational advice.</p>

Table 6: Interventions proposed for structure's criticalities (C_S) as defined in Table 3.

CRITICALITY	INTERVENTIONS
C_S type Z	<p>z1) Verify organizational times and conditions for the schedulation's adequacy with the structure of the course and the expected complexity;</p> <p>z2) verify the incidence of technical factors.</p>

Table 7: Interventions proposed for structure's criticalities (C_{PE}) defined in Table 4.

CRITICALITY	INTERVENTIONS
C_{PE} type A	a1) Acting on people by remembering tips for effective and productive study. Refer to the tutor for personalized advice.
C_{PE} type B	b1) Identify lagging users, and check their actual completion status and the sub-indexes.
C_{PE} type C	c1) Identify lagging users, and check their actual completion status and the sub-indexes.
C_{PA} type X	x1) The course is over. Among the course completed participants, more than 10% have an index of less than 70.

4. Visualizations and evaluation of critical issues in LearnalyzeR

In this section, we report examples of some courses in which, through the features of LearnalyzeR, it is possible to view and investigate the reasons that lead to the detection of critical issues with the criteria previously described. Using a medical paraphrase, we want to remember how critical algorithms identify symptoms. At the same time, through LearnalyzeR's visualizations of MIP, sub-indices, and variables that compose them, we can understand the causes (diagnosis) and prepare targeted tutoring actions (therapy). As our customers asked, in the following examples, we will show the logic of analysis passing from one course to another without fully showing the critical issues of one or more courses.

The first step in LearnalyzeR is the visualization of the median value of the Macro Index of Performance (MIP). While it can assume different states, we synthesized it using a colored dot over the median value using the following chromatic categorization (Figure 2):

- Gray: not calculable;
- Red: inadequate performance with many critical elements (< 50);



Figure 2: Median and IQR of the MIP for a course.

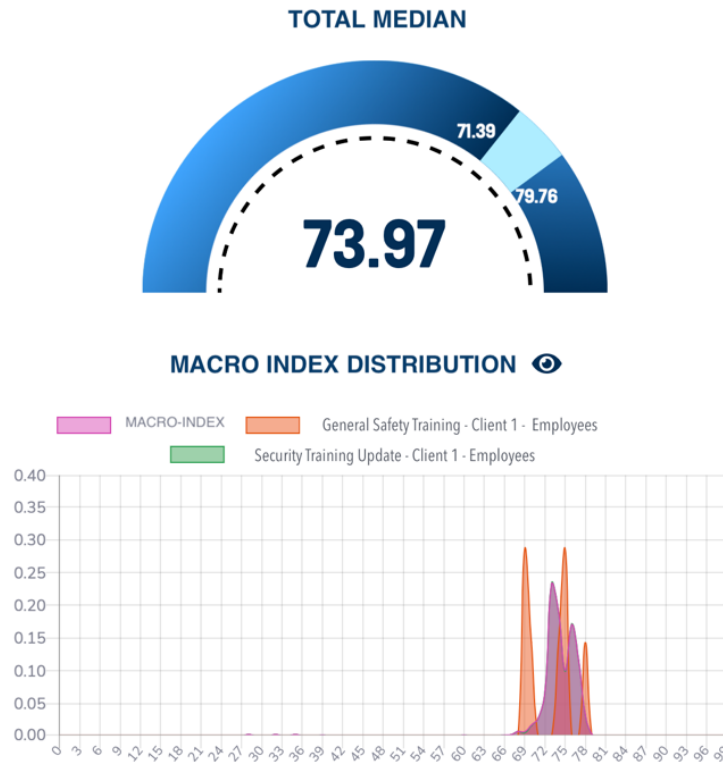


Figure 3: Safety at work courses, population 1

- **Orange:** quite adequate performance with few critical elements (≥ 50 and < 70);
- **Yellow:** good performance with areas for improvement (≥ 70 and < 80);
- **Green:** excellent performance (≥ 80).

In the rest of the section, we will report some examples from a public administration. The intent is to show how the Learnalizer framework (MIP, sub-indexes, and visualizations) matches the need to find the causes of course criticalities.

4.1. Examples on Public Administration Customer

For an Italian Public Administration, we gave two series of courses over time, Digital skills and Safety at work, on two different participant population types. Here are some compelling examples.

Example 1 comparison between different users population on the same courses

Here we will show the results for two populations on two courses (General Safety Training and Security Training Update).

We deduced a difference in performances between population 1 (Median MIP around 74, Figure 3) and population 2 (Median MIP around 65, Figure 4)

We observed a more significant variability of study behavior in population 2; there is a long queue on

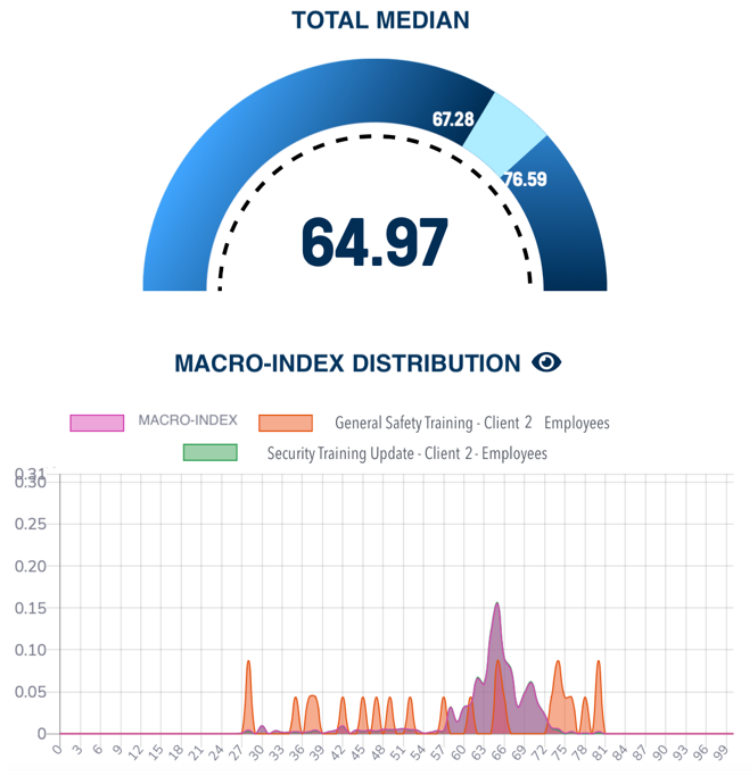


Figure 4: Safety at work courses, population 2

the left (towards the low values of the MIP up to a value of 27) for both Safety courses (General Training and Updating). Instead, population 1 has a homogeneous behavior in the MIP range between 69 and 78.

Example 2 – Detailed Analysis of a course criticality We analyzed the MIP and the sub-index in detail, observing the variables that make up the individual indicators. As said, the course structure has an average complexity because, despite having a high methodological complexity (value for variable methodological complexity is 100, the maximum), there is no test, and registration is not mandatory. The running course presents participation criticality of type d, and the sub-index Results have a median 50. In the case under study, this indicator directly gives the value at course completion while there is no test and no badge.

Both participation criticality and similar behavior between the running and concluded courses lead the tutor to undertake engagement actions. Users have good IT adequacy, but there is a problem with open sessions (variable open sessions is 60). The combination of the latter information with the structure criticality highlights to the tutor the non-effectiveness of using the time of use/study. The pace of study, although adequate, presents problems. Users have an adequate time of use, but effective days of study and the number of accesses are deficient.

According to forecasts, 95% of users are hard workers (final MIP predicted between 70.6 and 85.8, 95% C.I.). Still, comparing this running course with the concluded, we expect this forecast to change over time. We expect that there will be a dropout at 50% of course completion. Users in the first half of the course performed well despite problems in the distribution of study time: open sessions and low values of the actual study days variable. Macro-index forecasts are above the threshold, but all users do not go beyond 50% of the completion of both running and concluding courses.

4.2. Examples from analysis of didactical units

We need to observe the participants' behavior using individual didactical units (ud) to analyze the courses that present critical structure and the problem encountered in example 3 of Section 4.1. In

particular, we focused on the number of accesses in the individual ud and the time of use per ud. This analysis, combined with the knowledge of the course structure, allowed us to understand any critical issues in the individual ud:

- high or no number of accesses
- high or low time of use
- high population behavior variability in the number of accesses and the use time.

5. Conclusions

The systematic use of a criticality detecting system ensures the application of a regular and rigorous procedure for analyzing behaviors and results. It effectively monitors many courses and participants, even in different training courses for methods and durations. These measures also make it possible to effectively modulate data-driven tutoring systems, effectively analyzing the impacts in terms of effectiveness on the indicator. Adopting a statistical approach also means exploiting the information capital present in historical data to enhance the predictive system. Through the predictability and systematic analysis of the macro-index, sub-indices, and criticalities, tutors can act selectively in the most at-risk situations. They can modulate their intervention between the different categories of problems, individualizing communication and generalizing it or delaying it if valuable and less critical. Therefore, data-based tutoring becomes an intervention system consistent with the training system, independent of the tutors' sensitivities, guaranteeing a systematic and fair approach for all participants. Likewise, adopting a data-driven intervention system allows tutors to evaluate their actions in terms of effectiveness, striving to identify the best solution for individual critical issues. The tutoring action, therefore, starts from the same analysis scheme, guaranteeing tutors intervention methods that specialize in how to communicate with the participant. Above all, on these interventions, it is possible to evaluate the effectiveness with an analysis of the indicators and post-intervention criticalities to determine the actions/tools/messages with more excellent added value and more functional for the different categories of problems. The positive impact of the tutoring action is generally little measured and widely delegated to the tutors' expertise, who need to reread their efforts in terms of effects to evolve the intervention schemes to be an accurate guide of learning.

Our approach allows us to glimpse future steps to take. Integration of new features in the predictive and classification system will be:

- Include the tutoring actions performed following the analysis in the macro-index track, enhancing both the actual tutoring actions and those of a systematic type to evaluate their impacts on subsequent results.
- Study the evaluation thresholds of the sub-indices to decode the critical areas better.
- Integrate into the evaluation of the macro-index with extra measures for the sub-category Results with:
 - Focus on the analysis of learning assessment;
 - Observed results/work performance;
- Improve the satisfaction measurement in the category didactic tutoring also through NLP analysis.

References

- [1] T. G. Reio, T. S. Rocco, D. H. Smith, E. Chang, A critique of kirkpatrick's evaluation model, *New Horizons in Adult Education and Human Resource Development* 29 (2017) 35–53.

- [2] D. Pellegrini, M. Santoro, S. Zuzzi, Learning analytics and governance of the digital learning process., in: teleXbe, 2021.
- [3] D. L. Kirkpatrick, Techniques for evaluating training programs, *Journal of the American Society of Training Directors* 13 (1959) 3–9.
- [4] D. L. Kirkpatrick, Techniques for evaluating training programs: part 2 – learning, *Journal of the American Society of Training Directors* 13 (1959) 21–26.
- [5] D. L. Kirkpatrick, Techniques for evaluating training programs: part 3 – behavior, *Journal of the American Society of Training Directors* 14 (1960) 13–18.
- [6] D. L. Kirkpatrick, Techniques for evaluating training programs: part 4 – results, *Journal of the American Society of Training Directors* 14 (1960) 28–32.
- [7] T. T. Baldwin, J. K. Ford, Transfer of training: a review and directions for future research, *Personnel Psychology* 41 (1988) 63–105. doi:10.1111/j.1744-6570.1988.tb00632.x.
- [8] R. A. Noe, Trainees' attributes and attitudes: Neglected influences on training effectiveness, *Academy of management review* 11 (1986) 736–749.
- [9] L. A. Burke, H. M. Hutchins, Training transfer: An integrative literature review, *Human Resource Development Review* 6 (2007) 263–296. doi:doi.org/10.1177/1534484307303035.
- [10] D. L. Kirkpatrick, Great ideas revisited: Revisiting kirkpatrick's four-level model, *Training and Development* 50 (1996) 54–58.
- [11] D. L. Kirkpatrick, J. Kirkpatrick, *Transferring learning to behavior: Using the four levels to improve performance*, Berrett-Koehler Publishers, 2005.
- [12] R. S. Kaplan, D. P. Norton, *The balanced scorecard*, 1996.
- [13] R. Catalano, D. Kirkpatrick, Evaluating training programs: The state of the art, *Training and Development Journal* 22 (1968) 2–9.
- [14] P. E. Kennedy, S. Y. Chyung, D. J. Winiecki, R. O. Brinkerhoff, Training professionals' usage and understanding of kirkpatrick's level 3 and level 4 evaluations, *International Journal of Training and Development* 18 (2014) 1–21. doi:https://doi.org/10.1111/ijtd.12023.
- [15] D. Eseryel, Approaches to evaluation of training: Theory and practice, *Journal of Educational Technology and Society* 5 (2002) 93–98. URL: <http://www.jstor.org/stable/jeductechsoci.5.2.93>.
- [16] E. Bartezzaghi, M. Guerci, M. Vinante, *La valutazione stakeholder-based della formazione continua*, Franco Angeli, Milano, 2010.
- [17] M. Coldwell, T. Simkins, Level models of continuing professional development evaluation: a grounded review and critique, *Professional Development in Education* 37 (2011) 143–157.
- [18] D. S. Bushnell, Input, process, output: A model for evaluating training, *Training and Development Journal* 44 (1990) 41–43.
- [19] R. O. Brinkerhoff, *Achieving results from training*, Jossey-Bass, San Francisco, CA, 1987.
- [20] R. O. Brinkerhoff, S. J. Gill, *The learning alliance*, Jossey-Bass, San Francisco, CA, 1994.
- [21] R. Swanson, *Analysis for improving performance: Tools for diagnosing organizations and documenting workplace expertise*, Berrett-Koehler Publishers, 2007.
- [22] F. E. Holton, S. S. Naquin, New metrics for employee development, *Performance Improvement Quarterly* 17 (2004) 56–80. doi:https://doi.org/10.1111/j.1937-8327.2004.tb00302.x.
- [23] E. F. Holton, R. Bates, W. E. A. Ruona, Development of a generalized learning transfer system inventory, *Human Resource Development Quarterly* 11 (2000) 333–360.