

Learning analytics and governance of the digital learning process

Daniela Pellegrini^a, Mario Santoro^{a,b} and Sara Zuzzi^a

^a*Piazza Copernico srl, via Francesco Gentile 135, 00173 Rome, Italy*

^b*Istituto per le Applicazioni del Calcolo "Mauro Picone", Italian National Research Council, via dei Taurini 19, 00185 Rome, Italy*

Abstract

E-learning platforms collect a large amount of data on online courses, which progressively evolve towards complex and increasingly heterogeneous educational models. It is then necessary to pass the training's verification as linked to the observable results (completion status and results of the test), in favor of a global representation of the phenomena that positively or negatively affect the user experience. For this reason, 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. The developed Macro Index (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 previous similar data, guiding the tutors in the differentiated intervention methods to support and facilitate learning.

Keywords

Learning analytics, adaptive learning, tutoring, lms

1. Introduction

In recent years, the main reason there has been a proliferation of learning analytics is the need to understand participants' behavior compared to a training offer that becomes increasingly articulated in content and tools. The increase in digital training activities, the increase in users' digital experience, the advanced features available in tools, and training proposals based increasingly on blended learning methods have created an ecosystem of training data in constant growth. At the same time, there has been the growth of computational models and systems of data analysis¹.

Proceedings of the First Workshop on Technology Enhanced Learning Environments for Blended Education (teleXbe2021), January 21-22, 2021, Foggia, Italy


EMAIL: dpellegrini@pcopernico.it (D. Pellegrini); m.santoro@iac.cnr.it (M. Santoro);

szuzzi@pcopernico.it (S. Zuzzi)


URL: <https://www.linkedin.com/in/pellegrinidaniela/> (D. Pellegrini); <https://github.com/santoroma>

(M. Santoro); <https://www.linkedin.com/in/sara-zuzzi-96212866/> (S. Zuzzi)

ORCID: 0000-0001-6626-9430 (M. Santoro)

 © 2021 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

¹In the 2000s, the theme of *data mining* (computer science sector that uses techniques such as decision trees, Bayesian networks, statistic algorithms) has spread. Then in the 10's, we have observed the dissemination of structured models for big data management and analysis.

Historically the adoption of international standards (AICC before and SCORM and IMS, up to xAPI) has allowed structuring the information to make possible the interoperability of the systems and the effective tracking of the didactic activities. Today, the LMS platforms welcome an increasing amount of data on users' social interactions, personal data, academic information [1], intended as the participant's educational history. In this scenario, the tracking data analysis offers the institutes and corporate training departments a perspective on the participants' behavior during the training. They can plan (thresholds, updates, tutoring on the content) and make organizational decisions on this data (extensions, subsequent courses, re-enrollments).

The recent pandemic has helped to stimulate further what was already visible in recent years. There has been a significant proliferation of teaching methodologies (gamification, video learning, micro-learning, and social learning) and delivery models (webinars, simulations, mobile learning). They, including blended ones, are increasingly able to guarantee the quality of learning, engagement, and content coverage.

Today, understanding the monitoring data in depth becomes significant in terms of the learning experience and not only a mere detection of the completion states or the final results. Whether it is a digital course, classroom, or blended learning, adopting models of data analysis, it is necessary to remember that data describe the training experience and that learning remains an individual process and a substantially not quantifiable information. What the data can offer is, therefore, the recording of the essential and observable elements of the training experience, without judgments or inferences on the quality of learning. In harmony with the definition coined at *the first International Conference on Learning Analytics and Knowledge* in 2011, "*Learning Analytics is the measurement, collection, analysis, and presentation of data on students and their contexts, for understanding and optimization of learning and the environments in which it takes place*"²

As defined by John P. Campbell, Peter B. DeBlois, and Diana G. Oblinger [2], analytics are "*data analysis techniques that combine large data sets, statistical techniques, and predictive modeling. They can be configured as the practice of extracting institutional data to generate actionable intelligence*". It follows that the purpose of adopting analytic tools, which in the last decade have tried to make the tracking data of LMS platforms increasingly understandable, is, therefore, to make effective study behaviors intelligible, both to the participant and to organizations. It is essential to encourage them and personalize the training experience, supporting and promoting virtuous behaviors and, according to Siemens [3], to predict and support learning.

There are several learning processes that only partially pass through LMS platforms, including the famous and recently popular 70:20:10 model [4], which assigns significant importance to informal learning in working contexts, as well as social learning. However, corporate training is still significantly characterized by the provision of content both in a frontal mode in the classroom or webinar (ILT - instructor-led training) and in a structured and homogeneous way through online courses (typically in self-study), often

²First International Conference on "Learning Analytics and Knowledge(LAK)", Banff, Alberta, February 27–March 1, 2011,<https://tekri.athabascau.ca/analytics/>.

subject to mandatory or reporting requirements.

It follows that the training provided in the company is typically formal and generally passes through the LMS platform and its essential organizational functions (enrollments and registration of attendance in the classroom or in the webinar), as well as access to digital training contents, as well as tracking of the participation and study. Therefore, the didactic need (especially for technical-professional contents) is that of homogeneity of the intervention among all the participants, completeness of information, and verification of learning, which often result in the request to use all the contents and pass the evaluation tests.

How do organizations use tracking data?

- Typically, the use is for *compliance*³, that is, to verify the training's correct performance to fulfill the training obligations deriving from the regulations or from internal professional development needs.
- Evaluation of organizational aspects to improve the training service.
- Verification of the effectiveness of the design model⁴.
- Continuous improvement in the management of support processes for participants.

The organizational reasons for data analysis reside in uses that are substantially exogenous to the user's actual course of study, who instead constitutes the first stakeholder in data analysis. To make fair use of the tracking data and make the data effectively "talking", it is necessary to have an analytical model capable of representing the phenomena without distortions (with a firm root in data statistics). The model also needs to introduce and capitalize on the sector expertise to enhance the data itself information potential. Combining data analysis with professional expertise allows to:

- equip with tools to compare experiences;
- identify both effectiveness and risk areas;
- outline the characteristics the training proposal use;
- model governance actions for process improvement;
- adopt predictive patterns as a tool of participant engagement and motivation;

³A risk of incorrect data use lies in this need, from the point of view of what Sewyn (2019) calls performativity. That is the consequence of monitoring, which implies, for the participant, the need to "accountability" activities. It can lead to modify their own behavior merely to follow the indicators, producing distortions in study habits.

⁴Attention must be paid to the risk of so-called "dataveillance" as a risk of surveillance through data since the purpose must lie in supporting the study rather than controlling. Furthermore, the use of data must be functional to learning and not feed further models (for example of teacher performance verification) or enter a sort of data economy ([5]) in which the data themselves acquire an economic value for third-party tools.

- discover predictive patterns to adapt user support systems to improve study performance.

The article describes a model for all training delivery methodologies, characterized by a data-driven approach but enhanced with professional expertise. The model studied allows measuring behavior patterns that are useful for describing different teaching experiences scenarios and represent the analytical basis for differentiating support and improvement actions in the teaching process. The model also allows guaranteeing a didactic and organizational ecosystem increasingly modeled on the positive experiences of the learning group. In summary, the statistical indexes studied have the purpose of:

- describe and discriminate scenarios of participant behavior (learning experience);
- adapt intervention models and rules on the individual (adaptive learning experience);
- ensure a global and systemic view of the experience (global learning experience).

In section 2, we will describe the statistical methods and the programming language chosen to build and develop the model; in section 3, we will briefly report a case study on a blended course; in section 4, we will discuss the existence of ethical implication on our model.

2. Materials and Methods

2.1. Methods

A synthetic statistical index is a function of a finite or infinite set of values. It synthesizes the data by reducing the complexity of the problem under study, selecting the most significant aspects starting from the information obtained from the e-learning courses platform's database.

In general, a single index concerns only one or few aspects of the overall complex phenomenon, so a composite index has been constructed as a statistical function of synthesis of the indexes, to capture and summarize the main characteristics of the data in the study. The composite index comes from the aggregation of several indicators and provides global information on the analyzed phenomenon. The indicators inherit the quantification of the phenomenon that they analyze, but not the unit of measure that it loses. This allows the comparison of indicators that deal with different phenomena, even very distant from each other. The performance index and the indicators that compose it (also composed of sub-indicators, which will then be specified) can be used concurrently. They help the experts to identify the status and criticalities for each course, edition, and participant.

The interaction of a group (panel) of individuals (experts) who actively discuss a complex problem by identifying all aspects of the phenomenon creates a virtuous group

communication process based on the decision-making process, leading to indicators selection. With the same method for each index, they have identified a set of independent variables, obtainable from the platform, that characterizes it. The variables must be independent of each other, not to overestimate a particular aspect of the phenomenon. Once the indicators and variables were defined, the weighted arithmetic mean was chosen as the aggregation function and the Analytic Hierarchy Process (AHP) (Saaty) to calculate weights.

AHP, a multi-criteria decision-making method, is a versatile decision-making tool that allows rational evaluation of the quantities identified by defining a ratio scale from the paired comparisons. The primary objective of the AHP method is to combine the opinion of a sample of experts (panels) by interviewing them through a simple questionnaire with pairs of comparisons. This method allows for understanding the opinions on a specific topic without complications for the interviewee. Through the elaboration of the answers, it allows the construction of the decision-making process of the weights. Identify an expert group (panel) of which we use the experience and unique insights to define a phenomenon's structure. In our case, we had a higher level of complexity because the experts represent different fields of origin concerning experience and role.

2.2. Materials

2.2.1. Experts' panel

The panel consisted of 20 experts from different fields: e-learning project managers, commercial accounts, tutors, instructional designers, course production managers, graphic designers, course programmers, LMS programmers.

2.2.2. Software and Libraries, Algorithms

We wrote all the algorithms used with the open-source software R [6]. In particular, the `ahpsurvey` package [7] was used for weight estimation via AHP, the `betareg` package [8] for Bayesian beta-regression, while we developed the calculation and display algorithms for all the rest of the model.

2.2.3. Aggregated Performance Index

The analysis of the participants' performance represents the expression of a synthetic value (independent from the didactic method used) useful to define the learning experience in its complexity. It adequately values the results achieved, the commitment applied, the possible deepening and participation concerning the course's characteristics, and the course's adequacy for the participant. We identified seven indicators that characterize the participants' performance at each edition of a course (macro index I) to enhance the different aspects of the e-learning experience through consultation among experts.

The seven indicators are:

- Results (I_R),

- Study pace (I_{RS}),
- Course structure (I_{SC}),
- Computer adequacy (I_{AI}),
- Community participation (I_{PC}),
- Educational tutoring (I_{TD}),
- Process tutoring (I_{TP}).

Each index consists of characterizing variables measured or derived from the platform data. Both the calculation of the variables and their aggregation lead to continuous or discrete quantities between 0 and 100 (normalization).

2.2.4. Results Index

The Results Index (I_R) is an extensive⁵ quantity. It defines the achievement of the expected objectives of the participant. This index consists of three variables measured or calculated from data:

- Course Completion Percentage: the overall course completion percentage.
- Final Test Score: the best final test score.
- Additional performance badges awarded.

2.2.5. Study Pace Index

The Study Pace Index (I_{RS}) is an intensive⁶ quantity. It defines the distribution of study activities of the participant in the time available. This index consists of four variables:

- Total number of accesses: these are the total number of accesses to the course greater than 30 seconds.
- Study time: the sum of times for all activities (wbt, video, simulation, other activities) greater than 30 seconds; it does not include the time dedicated to the final test because it is not a study time.
- Actual study days: the actual days of study. As fruition time, it does not include the day of the test.
- Number of test attempts: is the number of failed test attempts, used as penalizing rate.

⁵It increases (not decreases) with the participant's progress in the course edition.

⁶It can increase or decrease as the course progresses

2.2.6. Course Structure Index

The Course Structure Index (I_{SC}) is a constant quantity compared to the edition. It defines the characteristics of the course in terms of planned activities and operating rules. This index consists of 4 variables measured or calculated from the data of the platform:

- Final test availability: is a dichotomous quantity with YES/NOT category.
- Type of enrollment: dichotomous quantity
- Mastery score presence: numerical, indicating the value of the mastery score.
- Methodological complexity: a variable that brings together complexity (defined by a panel of experts) by activity type (classroom, wbt, blended) and design duration.

2.2.7. Computer Adequacy Index

The Computer Adequacy Index (I_{AI}) is an extensive quantity that defines the connecting devices' inefficiency and the low computer participant's literacy. This index consists of three variables:

- Number of wbt accesses with problems: the number of accesses with a duration shorter than 30 seconds.
- Number of critic sessions: number of accesses with a duration longer than 8 hours.
- Number of helpdesk alerts: tickets opened with the helpdesk service.

2.2.8. Community Participation Index

The Community Participation Index (I_{PC}) is an extensive quantity that defines the participant's level of interaction within the community. This indicator consists of two variables:

- Number of Passive Interactions: represents the number of contributions' reading.
- Number of Active Interactions: represents the number of comments, responses, likes, and shares.

2.2.9. Educational Tutoring Index

The Educational Tutoring Index (I_{TD}) is an extensive measure that quantify the level of support on the course's contents requested by the participants. This indicator consists of three variables:

- The number of requests sent to the tutor: only didactic messages sent to the tutor.
- The number of replies sent by the tutor: only didactic messages sent by the tutor.

- Participant feedback: the number of participant feedback messages to the tutor's messages

2.2.10. Process Tutoring Index

The Process Tutoring Index (I_{TP}) is an extensive measure that quantifies the level of support on organizational aspects (information, deadlines, extensions, requested activities, reminders) requested by the participants. This indicator consists of two variables:

- The number of messages sent by the tutor: only reminder messages sent by the tutor.
- The number of responses to tutor messages: only responses to tutor reminder messages.

2.3. Procedure for calculating the performance index and sub-indexes

The aggregate performance index (macro-index) and the seven indicators are weighted arithmetic mean of the indicators and associated variables, respectively. The weights are calculated with the AHP method by submitting two questionnaires to a panel of experts. The inconsistency was calculated for each expert, and a limit value (Cho) was defined to establish the comparison's robustness. If the inconsistency of the answers for the single expert is greater than 0.15, the single expert's answers for that comparison are not considered. The final aggregation weight for the single variable is a geometric mean of the consistent answers, weighted with the interviewees' role.

Roles:

- Globalist (G): individuals who have an overall view of the platform, courses, and variables.
- Specialist (S): individuals specialized in a sector, which changes according to one of the seven indicators treated.
- Not Specialist (n_S): individuals not specialized in the treated sector.

For each of the seven indicators, the respondents change position between group S and n_S depending on the role. Figure 1 shows the performance index structure and its weights (obtained by the AHP method) for each indicator.

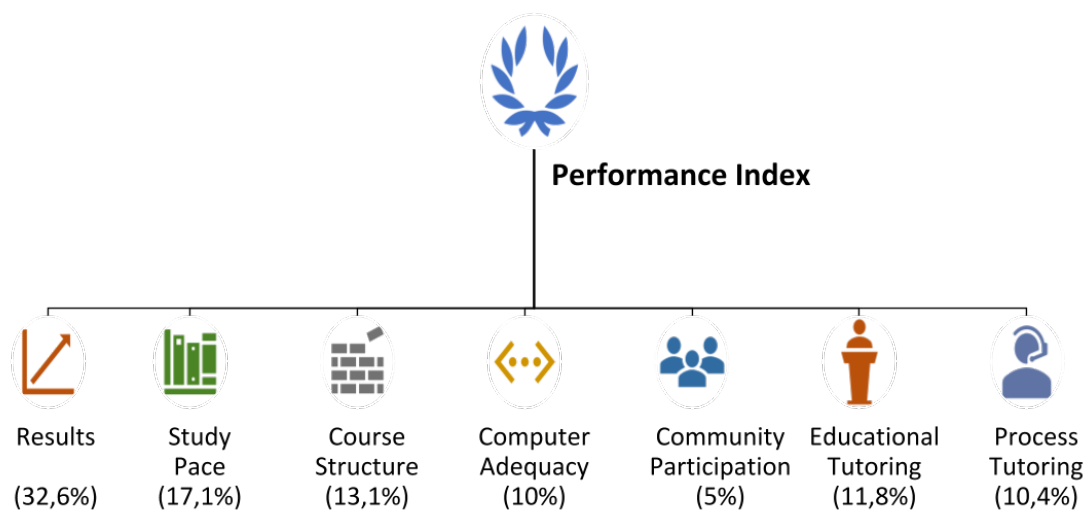


Figure 1: Performance index and sub-indexes. The values are the weights (in percent form) attributed to the indicators for the aggregation.

Table 1
Course Structure

Variable	Value
Final test	Yes
Course registration	Mandatory
Learning mastery minimum score	60 over 100
Classroom duration	9 hours
Online questionnaire duration	0.25 hours

2.4. Prediction Model

We built a prediction model for the macro-index at course completion (response variable) using macro-index evaluation at early stages (covariate). For a response variable such that, with a limited domain in $[0, 100]$, the suitable model is the so-called β -regression. This algorithm uses the β distribution as the probability density for the variable to be predicted since it is a very flexible function to represent variables in $[0, 1]$. As it appears clear, we developed an algorithm that maps the index in $[0, 1]$ and rescales the predictions in $[0, 100]$. We only considered the population that completes the course (then at the end of the course, the variable completion course is 100). We calculated the population's performance index when the course completion variable is $= \alpha^7$ (I_α) and used it as a covariate of the regression, whose response variable, as already mentioned, is the performance index $I = I_{100}$ at course completion.

3. Application to a blended course

This section will show the indexes calculation for a minimal blended course administered to a Piazza Copernico's customer. Given the client's privacy and anonymity choices, we will only show the index and sub-indexes results. However, this same general information will be sufficient to have a concise and exhaustive picture of the course's progress. In table 1, we show the course design structure:

Figure 2 shows the macro index distribution for the participants. The density clearly shows three behavior ranges: under 40, between 55 and 70, over 70.

Looking at figure 3, we can go more in details with the sub indexes⁸:

- Results sub-index has two spread ranges at the borders (under 25 and over 70). Detailed density is in figure 4.
- All participants have a Study Pace sub-index under 40.
- All participants have a Computer Adequacy greater than 70.

⁷As we will show in result usually $\alpha < .2$ that means completion $< 20\%$

⁸We do not show Community Participation, Educational Tutoring, and Process Tutoring because all the data are **NA** for this course. Since sub-index Course Structure has a unique value for all participants, then the density is infinite, and we will not show it in the graph.

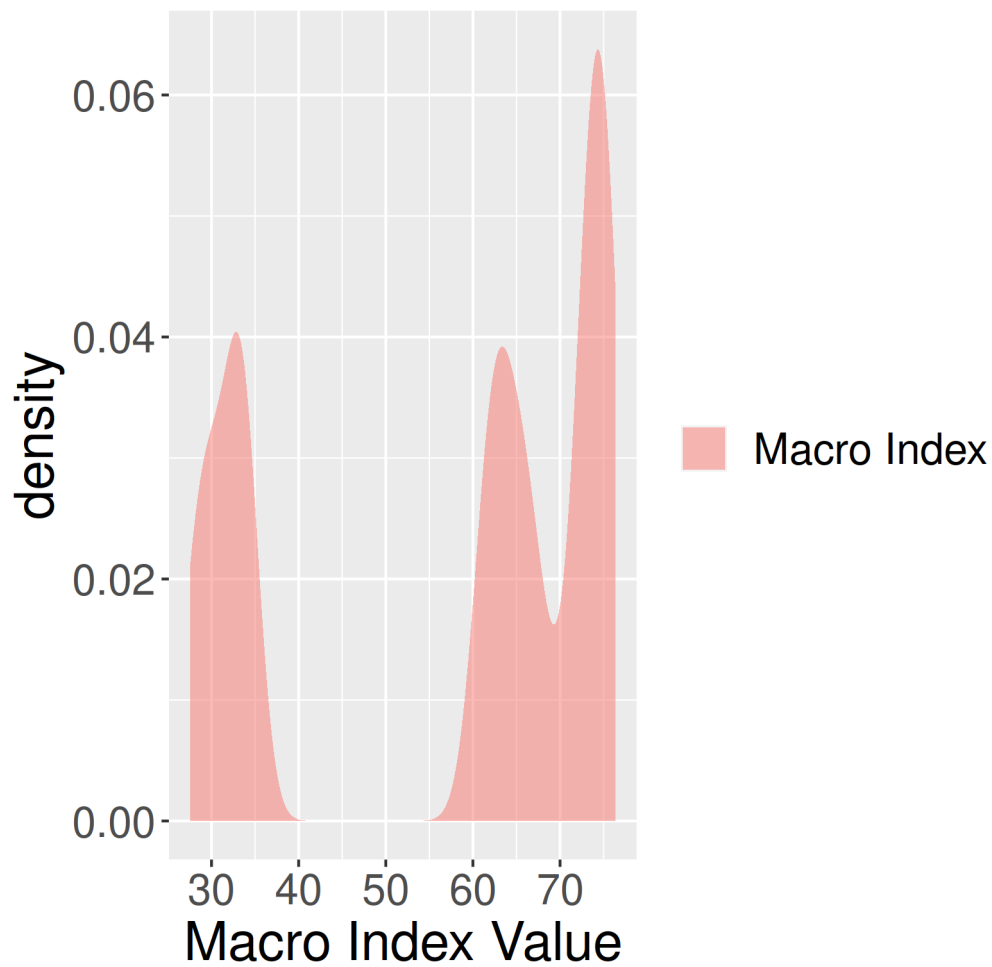


Figure 2: Computed distribution of the Macro Index values for all the participants.

Concerning the simple looking at results (figure 4), with two behaviors (brutally passing or not passing the test), using our indexes structures one can see more in details the effect of a lower, and maybe unexpected, study pace, resulting mainly in three groups of participants:

- Low study pace⁹ (under 30) and poor results (under 25).
- Low study pace (under 30) and good results (over 70).
- Not so low study pace (between 30 and 40) and good results (over 70).

⁹Respect to the expected/requested.

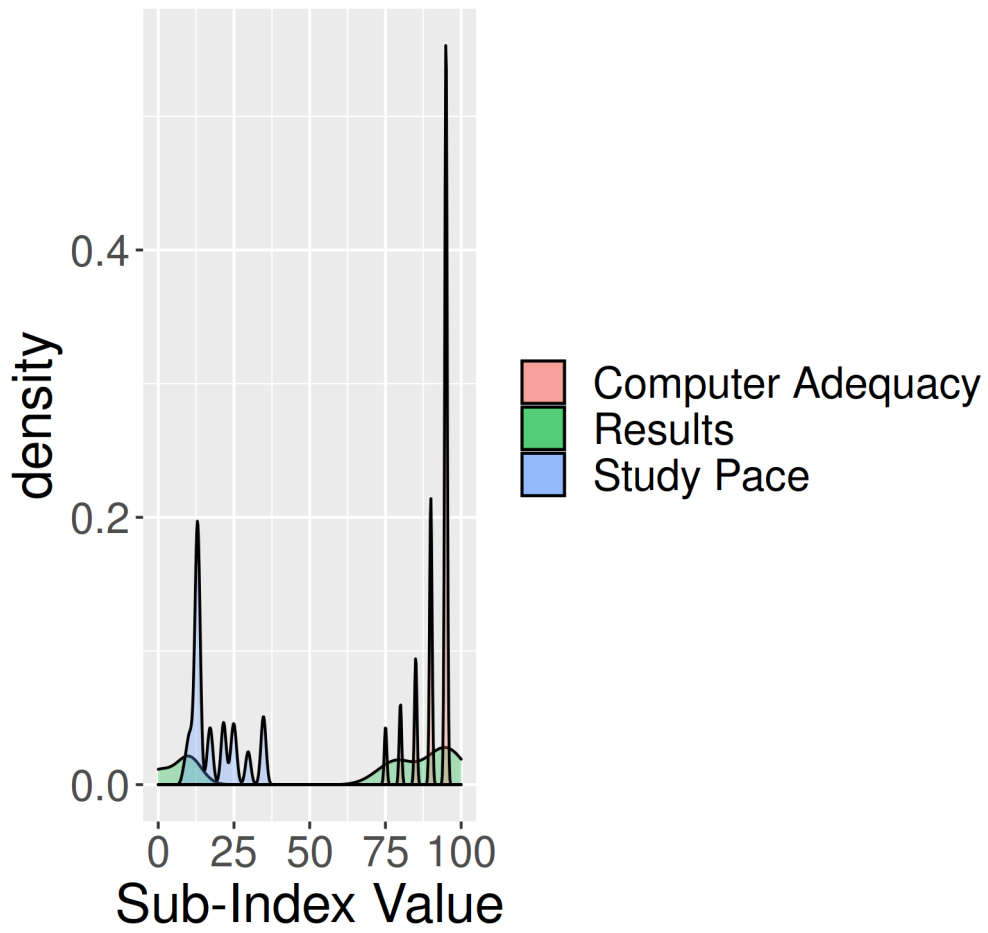


Figure 3: Computed distribution of the Computer Adequacy, Results, and Study Paces Index values for all the participants.

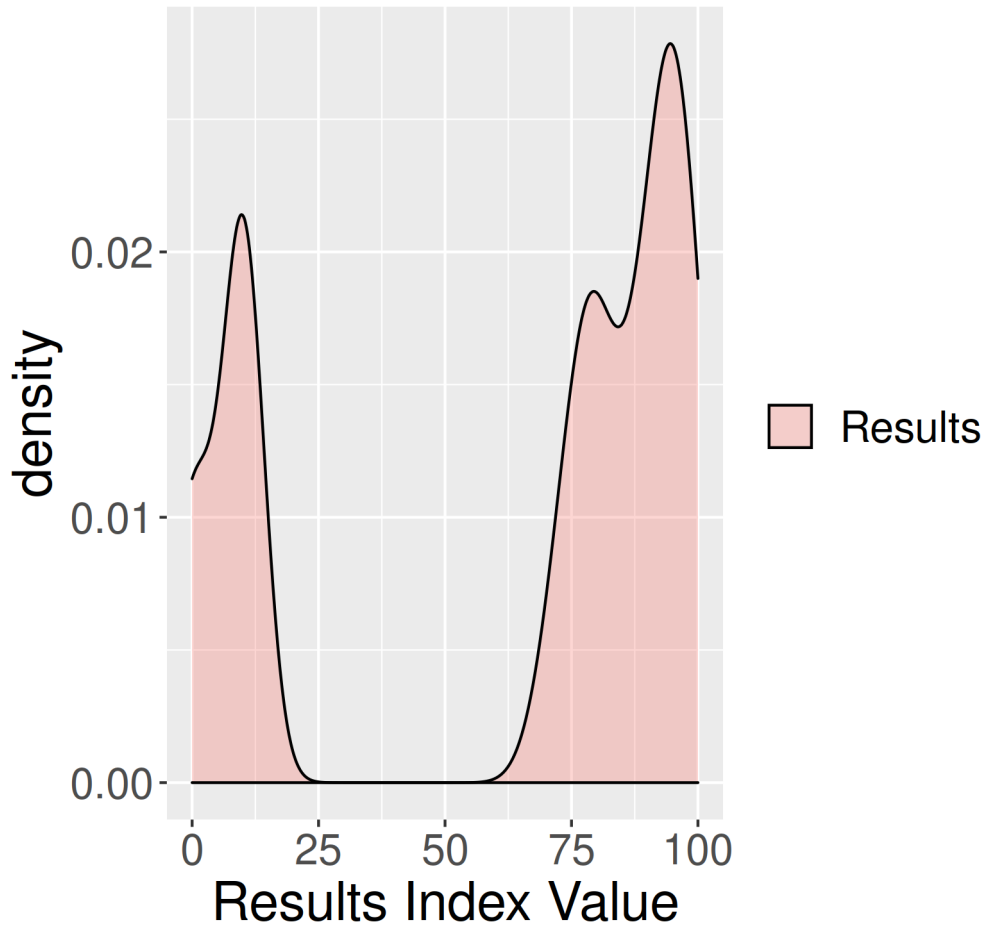


Figure 4: Detailed zoom on the computed distribution of the Results sub-Index values for all the participants. Keep attention to the scale: it is not the same as the previous figures!

4. From ethical learning indicators to the ethical use of data in HR

We considered the main criticalities and potential obstacles in adopting an analytical approach to user data on the LMS platform during the construction of the analysis model.

As highlighted by Selwyn [9], there may be various myths and risks connected to the theme of learning analytics, not least that of the **theoretical model**.

It was crucial in this experimental work on the Scorm compliant platform data to avoid any implicit bias in the analysis model.

The first risk prevention action concerned the need to *delimit the study's subject exclusively to the learning experience*, i.e. the participant's explicit behavior, without any specific assumption on individual learning processes. To separate experience from learning means observing the explicit dimensions of learning, avoiding the intention of "monitoring behavior concerning the expected ideal behavior" as correct and appropriate.

Furthermore, this approach allows relativizing the participant's experience, not limited to achieving results (e.g., completing the course or passing the test threshold). However, it incorporates the full expression of the study's enabling or disruptive variables (see sub-indexes in the model). Another possible bias recognized by Selwyn [9], might be due to the data's sampling and representativeness (with particular relevance of one group over the others). Concerning that, we choose to work on the overall dataset. The level of comparison was determined at the edition-course level and therefore stabilizing the variability linked to the organizational position, education, company seniority, gender, but fully reusing the company's organizational enrollment criteria.

Besides, we conducted each trial phase *to avoid a project asset's preponderance over others* (analysis model vs. statistical data vs. technical implementation models) through AHP techniques and continuous moments of confrontation of the work team. With the approach balancing, we tried to avoid the distortions due to the absolute confidence in the data, as *sincere technological idealism* [10], with the expectation of the technologies' effectiveness. The latter often generates a tunnel vision on the impacts in the educational and formative field or the so-called *tech-domain*: the confidence that the information's computer treatment can always guarantee superior and effective results. We developed the model, ensuring a structured analysis for the configuration and customization of the model. We fully exploited the specific company knowledge, the training context's characteristics, and the typical nature of the training provided. This process allows us to recognize the needs of context customization already in the start-up phase of the tool, identifying and enhancing users' needs and peculiarities.

However, there are still **ethical issues** related to data use in companies.

First of all, it is essential *to make the index data available to the individual user*, clarifying the analysis workflow (developed according to models for which researchers/designers take responsibility [11]) and its uses. This lead to transparency and *white box* applications [12], giving checking possibility over personal data and calculations [13], and especially promoting awareness of the reasons and methods in using personal data [14]. As

part of the tools created, this translates into giving access to both summary and detailed statistics, together with crucial *information on the characteristics of the analysis model and data use*.

It is also essential to explain the model's characteristics to all data users (especially those who consult them both at the individual and the global level). Since an index represents extreme synthesis statistics on the behavior (seen or expected), **the model sharing process** is essential to understand the nature of the data (also for the variables not included in the index), the calculation model, and the intrinsic meanings. All this process gives the ability to use the results as intended, in an ethical and relevant way.

5. Conclusions

The work had the purpose of studying a unified and global model for the analysis of training experiences. The project's specific objectives were:

- to define a single analytical metric to compare different synchronous and group training experiences (from the classroom to the webinar) or typically individual (from SCORM objects, exercises, videos, materials, assessments) or even blended courses. This metric represents a fundamental potential for training institutions to meta-evaluate the effectiveness of their offer and consolidate the most effective models based on the characteristics of the participants;
- to create an index capable of summarizing the complexity and the variables involved in a training project, mainly digital training, makes the user experience measurable and understandable beyond the results alone.

Through a complex study, which involved:

- the analysis of all the variables involved in corporate training encoded in the LMS platform,
- the analysis of data from significant company projects both numerically distributed and highly heterogeneous over time,

we built statistical indicators measuring learning experience both in progress and ex-post, and predictive indicators useful for setting up adaptive tutoring actions to mitigate the drop out risk.

Our work makes learning tracking data more understandable. It then permits to overcome the practices that look at the satisfaction questionnaire and the course completion as indicators of training effectiveness. The same tracking data can become motivational levers for:

- the individual, giving awareness of the behavior and forecasts, allowing him to improve where necessary,

- the learning support service, giving tools to set up tutoring strategies oriented to the participant's real needs.

Adopting an analytical approach has made it possible to enhance all the data already tracked daily in the LMS platforms that only partially express their information potential.

The further step will be to statistically study the index behavior to find participants clusters to differentiate the tutoring customized actions and define a customizable model based on the training context's characteristics.

Other uses of data relating to training experiences concern:

- to redesign critical training paths in the delivery phase (identifying the critical areas of the course to make it more effective);
- to meta-evaluating the offered training by identifying the most compatible methods with the participants' needs, different training models for different contents, correctly assigning the course availability time, or finally defining in a targeted way thresholds for overcoming the courses;
- to identify external phenomena that influence the training provision. Using indexes as statistical variables gives us the possibility to identify time variability and precisely evaluate statistical paths (in the sense of variables value combinations) that significantly differ from aspected behaviour. Thus, it verifies any scenario phenomena that affect the training process, both in positive, as acceleration conditions, or harmful, as disturbing factors, effectively increasing the training process's governance.

Acknowledgments

We wish to thank the director, company board, and the entire Piazza Copernico team for their collaborative participation in the preliminary analysis activities that led to collecting the variables underlying the indexes and offering the research work the specificities of the various professionals involved. A special thanks to the team who worked with the researchers on all the project steps: Daniele Poce, Paola Di Marco, Serena Di Sanno, and Marcello Pucci.

References

- [1] R. Mazza, C. Milani, Gismo: a graphical interactive student monitoring tool for course management systems, 2004. The T.E.L.'04 Technology Enhanced Learning'04 International Conference. Milan, Italy.
- [2] P. B. D. John P. Campbell, D. G. Oblinger, Academic analytics: A new tool for a new era, *EDUCAUSE Review* 42 (2007) 40–57.
- [3] G. Siemens, What are learning analytics?, 2010. URL: <http://web.archive.org/web/20100831015646/http://www.elearnspace.org/blog/2010/08/25/what-are-learning-analytics/>.

- [4] M. M. Lombardo, R. W. Eichinger, *The Career Architect Development Planner*, 1st ed., Lominger, Minneapolis, 1996.
- [5] K. Iveson, *Our Digital Rights to the City*, Meatspace Press, Oxford, 2017, pp. 20–22.
- [6] R Core Team, *R: A language and environment for statistical computing*, 2019. URL: <https://www.R-project.org/>.
- [7] F. Cho, *ahpsurvey: Analytic hierarchy process for survey data*, 2019. URL: <https://CRAN.R-project.org/package=ahpsurvey>.
- [8] F. Cribari-Neto, A. Zeileis, Beta regression in r, *Journal of Statistical Software* 34 (2010) 1–24.
- [9] N. Selwyn, What’s the problem with learning analytics?, *Journal of Learning Analytics* 6 (2019) 11—19. doi:10.18608/jla.2019.63.3.
- [10] C. Sims, *Disruptive fixation: School reform and the pitfalls of techno-idealism*, 1st ed., Princeton University Press, Princeton, NJ, 2017.
- [11] A. Smith, *Franken-algorithms: The deadly consequences of unpredictable code*, 2018. URL: <http://www.theguardian.com/technology/2018/aug/29/coding-algorithms-frankenalgos-program-danger>.
- [12] O. Tene, J. Polonetsky, A theory of creepy: Technology, privacy, and shifting social norms, *Yale Journal of Law and Technology* 16 (2014) 59—102.
- [13] J. R. Reidenberg, F. Schaub, Achieving big data privacy in education, *Theory and Research in Education* 16 (2018) 263—279. doi:10.1177/1477878518805308.
- [14] P. N. Howard, *Is social media killing democracy? computational propaganda, algorithms, automation and public life*. inaugural lecture to the oxford internet institute, 2017. URL: <http://www.youtube.com/watch?v=J1kXdA61AQY>.