Je-LKS
PEER REVIEWED PAPERS
LEARNING ANALYTICS: FOR A DIALOGUE BETWEEN TEACHING PRACTICES AND EDUCATIONAL RESEARCH
Vol. 15, n.3, 2019
ISSN: 1826-6223 | eISSN: 1971-8829
Journal of e-Learning and Knowledge Society
The Italian e-Learning Association Journal

USE OF LEARNING ANALYTICS BETWEEN FORMATIVE AND SUMMATIVE ASSESSMENT

Carlo Palmiero[1], Luciano Cecconi

INVALSI-University of Modena and Reggio, University of Modena and Reggio
palmiero.carlo@gmail.com; luciano.cecconi@unimore.it

Keywords: Learning Analytics, formative/summative assessment

The field of study within which this work is placed is that of data produced within digital learning environments, a field of research now known as Learning Analytics (LA). In particular, the aim is to investigate the relationship between the standard psychometric properties of the test questions and the information obtained from the log files produced during its administration, on a large scale, by computer. The results of this type of survey can help to make visible the intersections between formative assessment and summative assessment and to renew, in this way, the evaluation practices of a rapidly expanding sector such as digital education.

[1] The opinions expressed are to be attributed to the author and do not engage the responsibility of the Institute to which they belong.

for citations:
Palmiero C., Cecconi L. (2019), *Use of Learning Analytics between formative and summative assessment,* Journal of e-Learning and Knowledge Society, v.15, n.3, 89-99. ISSN: 1826-6223, e-ISSN:1971-8829
DOI: 10.20368/1971-8829/1135019

1 Process data: a new research frontier

The use of computers, or other digital tools (tablets, smartphones, etc.), to conduct large-scale assessment tests offers new opportunities, including for educational research, to acquire otherwise inaccessible information on how students learn. Changing the tool with which students carry out assessment tests, from paper to digital supports, has implications not only on the technological level but also on that of knowledge of the cognitive processes underlying learning.

The administration of large-scale computer-based testing (hereinafter referred to as CBT) makes it possible to collect information that cannot be retrieved when evidence is administered in paper form. In principle, CBT mode allows all interactions between the respondent and the testing platform to be recorded in so-called log files (LFs). This information, called process data literature (PD), allows to study the processes that lead the student to provide a certain response and, therefore, more generally can contribute to the observation of learning processes implemented by the respondent. They allow, for example, to trace the ways in which a student relates to a task (time of reading the task, time elapsing between reading and the first or last interaction with the task, number of attempts made to solve the task, etc.), thus providing much information on different cognitive styles and approaches to the task. It is clear that there has been a change of perspective with respect to the traditional evaluations, which are more focused on the observation of the final outcome than on the process that determined that outcome. PDs therefore become one of the most important subjects of study in a discipline that has only become internationally established in the last decade and is known as Learning Analytics (LA). The results produced so far by this new field of research have, on the one hand, contributed to raising the awareness of designers, teachers and managers of what is happening within digital learning environments and, on the other hand, have made more evident the need to involve the world of pedagogy in this field of research (Lang *et al*., 2017).

The study of PD allows to identify proximity variables (proxy) able to provide information on the respondent's motivation, his involvement in the task (engagement), his perseverance, etc.. Thanks to this information and its analysis, it is possible to deepen the knowledge of so-called soft skills, the importance of which is widely shared, but on whose observation and measurement methods many open and controversial issues still need to be clarified (Heckman *et al*., 2017). In fact, PDs can provide proxies for the respondent's behaviour styles in dealing with the proposed task. PDs can provide indications, albeit partial, of the processes activated and the character traits mobilized to reach a solution.

In this perspective, the use of PD makes the boundary between formative and summative evaluation less clear since the evaluation process, whatever it may be, tends to become a process of continuous observation and thus provides feedback that is beneficial to both students and teachers.

However, the interpretation of LFs is not easy, both from a technical and, above all, from a theoretical point of view. At present, LFs are often structured around the technical characteristics of the platform used to deliver CBT tests and are not designed, engineered or developed as an integral part of the assessment action. Although, over the last decade, LA applications have taken very important steps, it is crucial to define theoretical reference frameworks that can give them an adequate systematic approach, thus making them functional to the cognitive instances on learning processes and on the factors that can produce the rise.

PDs can provide a lot of information relevant for both formative and summative assessment; in fact, they allow a new point of observation on student behaviour, whose importance was long known in the literature (Bunderson *et al.*, 1989), but which current technological developments finally make accessible. Until now, the attention of teachers and researchers has been mainly, if not exclusively, focused on summative assessment, i.e. on the product. Today, the opportunities offered by the PDs to acquire information in real time on all the interactions that the student establishes not only with the evaluation tests but also with all the didactic activities (e-activities) that take place through the learning management systems (LMS), are an additional tool in the hands of those who want to start a process of profound renewal of the assessment (Hill *et al.* 2014; DiCerbo *et al.*, 2014), a renewal that places at the centre of the assessment actions the process-oriented dimension of the learning paths. A dimension that is much more congenial to the assessment than those complex skills that can hardly be detected and assessed only with the methods and tools of product assessment (summative assessment). The same summative assessment of complex products needs much more information to integrate those found with traditional assessment tools (tests, standardized tests, etc.), so a different interaction with the formative assessment may be useful.

2 The log files: use and potentiality

The CBT mode of large-scale measurement testing, on which this contribution is focused, is increasingly becoming the reference standard in more advanced countries (Parshall *et al.*, 2002). Since 2018, Italy has been one of the most innovative examples both from the technological point of view and in terms of the number of students involved (INVALSI, 2018). The INVALSI tests of the secondary school are in fact administered to all the

students of the third class of the first grade secondary school, of the second grade secondary school and of the second grade secondary school. According to a modern CBT design, students carry out item bank tests based on rigorous psychometric principles that allow them to obtain results for each student on longitudinally comparable Rasch metrics (1980). CBT tests increase the efficiency of administration, the ability to monitor the entire process and reduce errors in the design, implementation and correction phases.

The most important aspect of CBT administration is that it enables the collection of process data (PD), which can acquire information on any interaction between the respondent and the platform that delivers the test (Greiff *et al.*, 2015, p. 92). This is an opportunity that can also prove very useful in the measurement of complex skills, such as soft skills, increasingly cited in schools and universities. However, the problems related to the development and measurement of soft skills are very relevant, because of their obvious intersection with areas from which the school has retracted over the past decades. Several soft skills are related to the sphere of character or aspects very linked to cultural, political, religious visions, from which the mass school in complex and heterogeneous societies has deliberately withdrawn. Some examples are conscientiousness, persistence, openness towards the other. However, their indirect observation using PDs, such as problem solving, can open up very interesting research and application scenarios.

The availability of PDs allows the point of observation to be shifted from the respondent's outcome with respect to a task (the response to a stimulus, open or closed) to the entire process leading to the production of that outcome. The positive aspects of this change of perspective are evident, especially if we consider this change from the point of view of those who observe the individual behaviour of the individual respondent to identify where their learning difficulties lie and to prepare, therefore, appropriate compensation interventions.

PD-based research is still in its infancy, both from the technological point of view and from the more properly theoretical-methodological one. Currently, LFs containing PDs are almost exclusively defined on the basis of the technical characteristics of the platform delivering the CBT tests. This leads to considerable difficulties in their use, but especially in their analysis and interpretation. The development of a general theory that can affect the structuring of PDs, but above all that defines what is important to observe and in what perspective, is almost completely lacking. In this respect, pedagogy, which sees the teaching-learning process as the centre of its interests, can make a very important contribution to the development of LA research.

Even today, PDs are still very much linked to the specific characteristics of the questions to which they refer. This is both an advantage and a limit.

The positive aspect is certainly represented by the fact that the data provide a rather precise and articulated representation of the respondent's behaviour with respect to a given task. However, the information drawn from PDs so linked to a specific type of question is difficult to generalize and compare.

From what has been briefly explained, it is clear that it is appropriate to move from LF, which are, in fact, a sort of technological sub-product, to LF, which become an integral part of the design of the assessment test, whether formative or summative. In this way it is possible to improve the accuracy of the outcome measures by introducing innovative question typologies, after having actually verified their impact on the respondent's behaviour. But above all, it is possible to assess the behaviour of the person who faces a task, observing the processes followed and the resources mobilised in the situation, thus also opening up the possibility of observing competences not strictly related to the construct being measured.

These new scenarios may also trigger new paths of research in other scientific fields, which are crucial for a full understanding of any evaluation process, on a large scale or on a smaller scale. Think, first of all, of educational research and its possible contribution to the definition of the conceptual framework of reference, which is essential to guide the basic decision-making processes for the analysis and exploitation of results. But also in the psychometric field, which can be used of PD to define multidimensional models and not simply multivariate for the definition of the results of a test (Ercikan & Olivieri, 2016, p. 310).

2.1 The value of the log files

It has been said that LFs are generated by the respondent's interaction with the platform that manages the CBT test. The data contained in them, the so called PD, are assimilable to the big data and therefore they are easy to process through the typical methods of the Artificial Intelligence (AI). This opens the way to many opportunities, but also to significant risks. A predominantly exploratory and non-confirming approach, if on the one hand it can open up new horizons of knowledge that cannot be identified a priori, on the other hand it sees a reduction in the methodological, but also ethical control that can be exercised by the pedagogical theories that are involved (Harari, 2018).

The typical structure that LFs have today makes them difficult to understand without the intervention of specific software to facilitate their inspection and analysis. On this aspect, the LA scientific community is producing several open source resources, such as platforms to deliver evidence capable of collecting PDs. However, this is not enough. It is increasingly important that the structure of the LF and the nature of the PDs are defined in the design phase of an

assessment test.

PDs can provide a set of information that goes far beyond the final outcome of the respondent's activity with respect to a task (e.g., task reading time, time between reading and the first or last interaction with the task, number of attempts made to solve the task). They would allow to better know the processes that are activated during the performance of the task, to improve the quality of the questions, to customize the analysis of the answers. In addition, PDs would allow to identify more appropriate procedures to prevent unwanted phenomena such as cheating[1] and data fabrication[2], phenomena that are increasingly relevant in the use of CBT tests.

3 Time indicators to investigate respondent behaviour

The first large-scale uses of PDs mainly concerned time indicators. It is still a limited use and, in some ways, only exploratory, but still able to make much more rich and informative the use of PD in the near future.

In this contribution, reference is made to data from the Programme for International Assessment of Adult Competencies (PIAAC) with the aim of showing some of the potential of PDs.

The PIAAC study is conducted periodically by the OECD on a sample basis in the adult population aged between 16 and 64 years. The research is carried out in a sample of about 5,000 adults in each participating country. Data are freely available from OECD website. The aim of the research is to verify the literacy and numeracy skills of adults of working age. The questionnaire consists of literacy and numeracy tasks. The survey is carried out in the presence of a specially trained interviewer. The interviewer has the responsibility of assessing whether the interviewee has basic knowledge in the use of computers. If the verification is successful, the rest of the survey is done by computer, otherwise it continues in paper and pencil format. The PIAAC research is particularly interesting for the purposes of this paper. As these are basic skills of an adult population no longer in school, the strategies and processes activated by the respondents to address the tasks are even more relevant to interpret the answers provided.

PIAAC provides three time indicators:

1. total time spent on an item (time on task);
2. reaction time to the item, i.e. the time between the item being displayed and the respondent's first action on the platform (time to first interaction);
3. time elapsed between the last action on the item and the final confirmation

[1] A set of fraudulent behaviours of the student or teacher in which the correct answer is given or drawn from illegal sources or following the suggestion or direct intervention of the teacher.

[2] Automatic data/response production process capable of delivering hundreds of thousands of tests in a short time.

of the response (time since last action).

The interpretation of time indicators may not be easy and, above all, may lead to very different conclusions. Once again the need emerges for the definition of a theoretical framework of reference that allows the introduction of a confirming dimension in the analysis and interpretation of PD.

Another very important aspect in which PDs can make a significant contribution is the study of missing data. They can represent the epiphenomenon of cognitive and motivational aspects of extreme informative and interpretative relevance. PDs allow you to begin to shed light on how and, potentially, why this missing data is produced.

The PIAAC data show different behaviours with respect to the three above-mentioned indicators based on the level of test result, with respect to the country of origin and the personal background of the respondent. These first and provisional indications seem to encourage the deepening of the study of PD to identify the implementation of different processes, depending on the characteristics of the respondent, for the performance of the same task. Enormous study opportunities would open up for the promotion of positive actions, based on solid and relevant empirical evidence, aimed at improving learning levels. This is a field of research and action perfectly in line with one of the most important functions of LA, that of predicting difficult situations or situations at risk of failure.

Of the three time indicators, the first (TOT) is the easiest to interpret and the most informative. With all due caution, TOT can be considered as a proxy for the respondent's commitment to the task they are facing. This indicator is produced by the interaction of several factors, the most important of which are: a) the level of competence of the respondent, b) the involvement and commitment of the respondent, c) the psychometric characteristics of the question, d) external events of various kinds (distractions, unforeseen events, etc.).

Figure 1 shows a strong variability in the distribution of TOTs between the different questions.

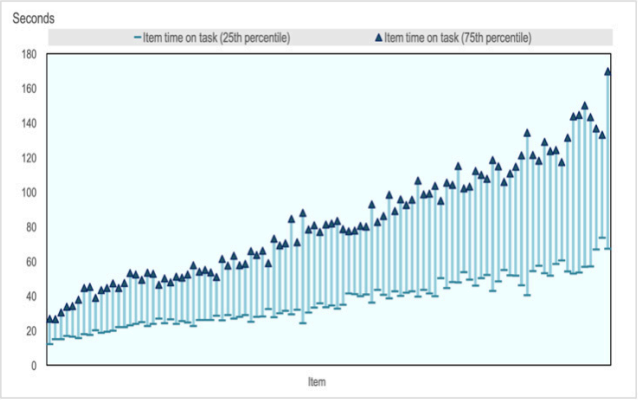

Fig. 1 - Distribution by item of TOT. Source: OECD (2017). PIAAC Log Files.

You can see that there is a large variation in TOT both within and between items. It is therefore clear that the study of the distribution of TOT opens up multiple lines of investigation that allow different assessments to be made on the characteristics of respondents.

Even more interesting is to evaluate the interaction between the median value of TOT of each question according to the type of answer (correct, wrong and missing).

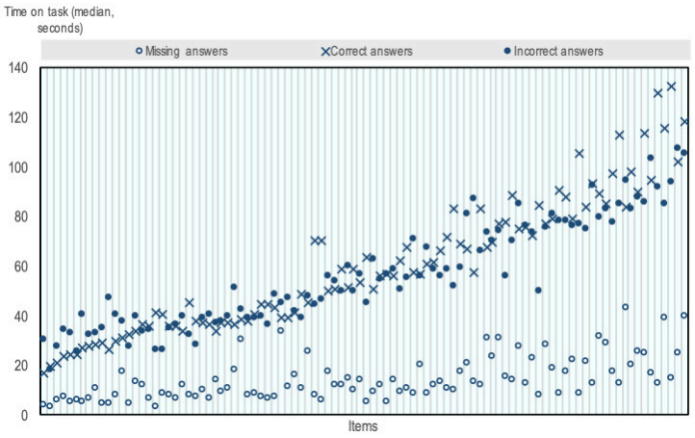

Fig. 2 - Distribution by item[3] and typology of response of TOT. Source: OECD (2017). PIAAC Log Files.

[3] Numeracy and literacy items. The lower end represents the 25th percentile in the TOT distribution aggregating the data of all the countries that performed the PIAAC CBT test and the upper end the 75th percentile.

Figure 2 shows that for each item the respondent providing the correct answer does not stay on the task for a very different time than the respondent providing the wrong answer. This seems to be further proof of the fact that TOT is strongly linked to the characteristic of the item, even more than to that of the respondent. On the other hand, a much lower variability than TOT can be observed in the case of missing answers. This seems to indicate that the decision whether or not to answer is more related to the respondent's characteristics than to those of the question.

Concluding remarks

The technological transformations of recent decades have been so profound that they have ended up influencing the very organisation of our societies, much more than other changes, such as cultural and political changes, have influenced them. However, some authoritative scholars have argued that digital technologies have not produced, in the educational field, those "mega-movers" that instead have produced in other sectors such as medicine, telecommunications, transport or the entertainment industry (Papert, 1993) and that investments in technologies have not been aimed at changing the educational system but only to increase sales of products (Laurillard, 2012; Fullan *et al*., 2013, p. 310). As in all transition phases, even today it happens that while some solutions are proving less and less suitable to solve the problems, there are still no alternative solutions or substitutes for the former. In short, a deadlock between what is no longer appropriate and what is not yet fully available, not by chance called "the swamp" (Fullan *et al.*, 2013, *Ibidem*). Evaluation, within the educational processes, is one of the areas most affected by this stalemate. Many people have seen the development of digital technologies as an important opportunity to renew the evaluation deeply. In 2014, Hill and Barber said: "Next-generation learning systems, however, will create an explosion in data because they track learning and teaching at the individual student and lesson level every day in order to personalise and thus optimise learning. In an online world with intelligent software and a range of devices that facilitate unobtrusive classroom data collection in real time, the big challenges will lie not so much in obtaining data but in managing it and protecting privacy while turning it into powerful knowledge, something that data warehouses built just a few years ago were never designed to support" (Hill *et al.*, 2014, p. 55).

Well, the next-generation learning systems is among us, the explosion of data is in full swing (Hill & Barber, 2014) the real challenge is to be able to manage the huge amount of data made available by digital technologies and ask the right questions. Another important challenge is to be able to give the right

importance and the most suitable tools to the evaluation of the process, in a system that has always favoured the evaluation of the product, the final results.

Education and training are affected by changes that need to be addressed in a constructive and balanced way, resisting both past temptations and impulses that, influenced by a kind of technological determinism, emphasize the innovative capabilities of technologies. The objective must be to restore to the educational system that centrality which is progressively losing, using all the resources made available by technological innovations. Artificial intelligence and, in particular, LA will increasingly be the subject of research in the field of education, the results of which may become important information supports for policy makers and for those responsible, at different levels, for educational institutions.

In this contribution we have tried to highlight some aspects related to LA and their potential. If LA, and therefore also PD, are used within a framework that is also pedagogically connoted, then they can become an integral part of the evaluation processes of education systems, contributing significantly to improvement, both at the micro level (the classroom, the school) and at the macro level (the whole system). In this perspective it is therefore possible to make the formative evaluation and the summative evaluation interact, determining a reciprocal and positive contamination, for the benefit of the learner, but also of those who plan the formative actions at various levels.

LA finally seems to offer a concrete way to implement what has always been hoped for, but which has so far proved almost impossible to implement on a large scale. LA can in fact be useful for personalisation, for a more careful evaluation of processes, understood not as a scale down of the importance of goals, but as a decisive factor in them.

All this will be possible, however, if the LAs do not follow only what in the statistical sciences is a clear exploratory approach, i.e. to search in the data for criteria and principles that are not desired or can be established on the basis of a theoretical framework. There is a need for in-depth methodological and theoretical reflection, which certainly finds its natural, though not exclusive, place in the sciences of education.

REFERENCES

Bunderson, C., Ynouye, D., Olsen, J., (1989), *The four generations of computerized educational measurement, in Educational Measurement*, 3rd ed., American Council on Education.

DiCerbo, K. E. and J. T. Behrens (2014), *Impacts of the Digital Ocean on Education*, London. Pearson. Available at https://research.pearson.com/content/

plc/prkc/uk/open-ideas/en/articles/a-tidal-wave-of-data/_jcr_content/par/articledownloadcompo/file.res/3897.Digital_Ocean_web.pdf

Ercikan, K., Olivieri, M.E., (1989), In search of validity evidence in support of the interpretation and use of assessments of complex constructs. Discussion of research on assessing 21st century skills, in Applied Measurement in Education, 29 (4), 310-318.

Ercikan, K., Olivieri, M.E., (1989), In search of validity evidence in support of the interpretation and use of assessments of complex constructs. Discussion of research on assessing 21st century skills, in Applied Measurement in Education, 29 (4), 310-318.

Greiff, S., Wüstenberg, S., and Avvisati, F. (2015), Computer-generated log-file analyses as a window into students' minds? A showcase study based on the PISA 2012 assessment of problem solving, Computers & Education, Vol. 91, pp. 92-105

INVALSI (2018), Rapporto prove INVALSI 2018, in https://invalsi-areaprove.cineca.it/docs/2018/Rapporto_prove_INVALSI_2018.pdf

Harari, Y., N. (2018), *21 Lessons for the 21st Century*, Random UK.

Heckman, J.J., Humphries, J. E., Kautz, T. (Eds.) (2015), *The Myth of Achievement Tests*, University of Chicago Press.

Lang, C., Siemens, G. Wise, A., Gašević, A. (Eds.) (2017), *Handbook of Learning Analytics*, Society for Learning Analytics Research.

Hill P., Barber M. (2014), Preparing for a Reanaissance in Assessment, Pearson. https://www.pearson.com/content/dam/one-dot-com/one-dot-com/global/Files/about-pearson/innovation/open-ideas/PreparingforaRenaissanceinAssessment.pdf

Laurillard, D. (2012). *Teaching as design science*. London: Routledge.

Michael Fullan and Katelyn Donnelly (2013), Alive in the Swamp. Assessing Digital Innovations in Education, Nesta-NewSchools, Report. https://www.nesta.org.uk/report/alive-in-the-swamp-assessing-digital-innovations-in-education/

OECD (2017), PIAAC data, https://www.oecd.org/skills/piaac/.

Papert, S. (1994), *The Children's Machine. Rethinking School in the Age of the Computer*, Basik Books.

Parshall C.G., Spray, J.A., Kalohn, J.C., & Davey, T. (2002), *Practical considerations in computer-based testing*. New York, NY: Springer

Rasch G. (1980), *Probabilistic models for some Intelligence and attainment tests*, Chicago, the University of Chicago Press.