DATA ANALYTICS EXPERIENCE: FINDING EQUILIBRIUM BETWEEN SURVEYS, EVALUATION AND MODELING

O. Sladek[1], B. Kamrlova[2], T. Zeman[3]

[1]KYBERTEC, s.r.o. (CZECH REPUBLIC)
[2]Atos IT Solutions and Services, s.r.o. (SLOVAKIA)
[3]Faculty of Electrical Engineering at CTU in Prague (CZECH REPUBLIC)

Abstract

The “new education era” is marked by new technologies applied in education. Because of the novelty of the content, there is a need for the new background concerning the research work in this field and, eventually, for a complete overview of the specific situation within a new educational platform with data analytics, including the management and analysis of feedback from students to teachers. The development in this field is driven by new technical approaches like virtual labs, gamification, mulsemedia, a new level of multimedia. This situation brings some new challenges and opportunities, eventually useful for our close future too. Almost any new view can be helpful in this promising research directions. To serve this purpose, we would like to present the survey, conducting a brief overview of experience with new educational platform and evaluation of its effectivity. Our survey also targets the gap between technical frameworks solution for learning and process evaluating its effectivity. Presented material also aims to review strengths and limitations of actually existing methods, from the point of view of the potential future directions for such a survey data analytics, with the data modeling based on surveys included. Last but not least explanatory power limitations determined by the number and/or distribution of respondents are described. To the best of our knowledge, this survey described selected effective techniques for data analyzing in this era of new e-learning platforms.

1 INTRODUCTION

The basis for this work was a part of the NEWTON project [1], concerning the new educational digital platform aiming to reinforce and improve the STEM subjects teaching/learning on all primary, secondary and tertiary level. Both digital platforms [2] and STEM education [3] attract more and more attention as well in the world of education [4] as of data analysis [4]. Our work deals with methods and tools of innovative technologies for modern forms of education. In particular, it creates experiments as interconnected existing teaching laboratories, a learning system platform, and new learning practices, including a comprehensive system for evaluating the effectiveness of learning processes.

As part of this project, we were confronted with the need to create a universally usable environment for preparing, collecting and evaluating learning information as part of a digital environment working with the LMS system being developed. At the same time, this approach was also tested on data from other systems.

2 METHODOLOGY

The overall approach included a sequence of development and use of individual parts:

2.1 Development of analytical toolbox for gathering information from learners and analytical evaluation

There are several research made about the innovative approach in STEM education using the most advanced technologies [3], [5], but it was hard to find the right way to catch every aspect of our research. The aim was to create a universal environment that would make it possible to obtain data in a transparent manner, in several forms, to store them and to work with them. To reach as large dataset as possible with the most relevant outputs, we had to move further from the content and get closer to the form or shape. Therefore, as a platform, it was chosen to develop a web environment with data storage in SQL database. The required functionality was the data input interface, the administration and analytics interface, and the interface for exchanging data with other systems. Other requirements that arose during the project and are interesting in this context are the necessary emphasis on the protection of personal data respondent incl. Anonymization in accordance with GDPR and then creating a general interface to databases so that some analytical work can be done outside the environment by connecting to tools in MS Excel, MATLAB etc.

For this reason, process modeling in UML was done at the beginning of the process to optimize DB table designs as well as the application's own interface.

There was also a process to prepare your own platform and verify the validity of the data contained therein incl. Processing test results pre-calculated on other systems.

2.2 Gathering information from subjects of learning processes

This process aimed to establish a good data acquisition methodology. To summarize their form, then codify it using generic XML and UML descriptions. The next assumption was to process all the data and verify their consistency, which proved to be a very difficult process to be described below.

As already mentioned, verification and development took place within the NEWTON project, where a fairly common course structure (pilot courses with hierarchical structure and country implementations) was established to validate different educational technologies (virtual and augmented reality, mulsemedia, FabLab, adapted learning etc.) being used to enhance the STEM subjects learning and learners ‘attitude to these subjects.

Quite traditional structure of questionnaires was defined for all learners participating in the pilot session/s like Demographic Questionnaires, Affective State Pre-questionnaires, Affective State Post-questionnaires, Usability Questionnaires. The second part of the similar questionnaires was focused on teachers. To complete questionnaires, there were two knowledge tests to be administered, pre- and post-course content. Interviews, focus groups and individual suggestions were to be included too.

To reach the best possible sample quality, more than 1000 learners and more than 100 teachers participated in the whole testing processes, divided in groups of various sizes from 5 to 140 learners, working on 32 small-scale or large-scale pilots. The variability of size and character of the sample was reinforced by the internationality (8 countries) and the age of learner’s span (10-25 years) as well as their educational level (lower and upper secondary classes, general and vocational secondary schools, university students).

Emoticons were used to facilitate the choice of answers, especially because of younger pilot courses participants. Distribution of questionnaires was prepared in electronical and printed form. Following the pre-use of questionnaires, it showed significant differences between the two forms – the e-form being much less carefully filled and, especially for the motivation questions, the suggestions for “other use” or “what did you like the best” et similar were rare. Therefore, the paper form was used for the project.

Learners were asked to put a cross in the cell corresponding to their choices Questionnaire was standardized and translated to the local language, but codification was respected, as well as the “emotional force” for the concerned questions in the attitude part of questionnaire. For anonymized identification of learner, the learner NEWTON single ID was associated to each questionnaire. As the standard research protocol requests, it was also up to all pilot leaders to enter all data in the NEWTON database in local form, scanned.

The same processes worked with the knowledge tests, although the extent of the whole project altogether with its internationality had slightly limited the data standardization – the content span of each course/pilot being completely different, there was no possibility to standardize the knowledge test as the whole. We had only managed to frame the duration (10-15 minutes each test) and number of questions (7-10). It was impossible to reach the common scoring (partial score/question and, therefore final result was different for each pilot) and evaluation as the classification rules/habits/norms are different for each country or even school.

All knowledge tests were to be scanned and assembled together with questionnaires.

The interviews were semi-guided and semi-opened with the given lines and basic keywords to be used. The same background was given for the focus groups. All interactions were registered and their transcription uploaded, classified by keywords.

The second reason for the paper questionnaires administration was a possibility of the immediate interview with the learner/learners/teacher. During several smaller pilots there was possible to discuss the subject both before and after the course what allowed researchers to pick the “live impression" of the pilot impact. This brings us three types of data: questionnaires, knowledge tests, verbal expressions.

2.3 Evaluation process

The evaluation process had four stages.

1. Data transcription, data cleaning and anonymization for each pilot, school and topic
2. Transcription of interviews, focus groups and learners ‘individual written verbal expressions
3. Analysis of the prescribed combination of questions from questionnaires
4. Overview of all possible indicators and "interesting points" following the data transcription

Each stage had its specificities and the evaluation was made on three levels – for each pilot separately, for each topic and for the whole dataset.

Each pilot leader was to find the most suitable way of data transcription/interpretation. This was the first issue halting the smooth process of evaluation. The standard data analysis processes would have required the standard approach in data transcription too, backed by standard data preparation and collection, guaranty of the final standard interpretation [6], [7].

The high variability of the whole group and data gathered from learners had eliminated this possibility. There were many invalid questionnaires as well as many copied answers visibly without any effort to really answer the questions. The post-assigned sub-groups following four selected parameters from questionnaires were too late attributed and, last but not least, did not allow the global analysis of the whole dataset, working only with the averages, without the ability to incorporate contexts and emerging findings and conclusions.

However, we had a possibility to compare the three levels of results using classical analytical tools. It was possible to have a general and detailed look over the huge datasets as well as the tiny individual pilot results, we could go through questionnaires and knowledge tests scoring growing or fading across the pilots or schools and checking our evaluations by verbal expressions.

3 RESULTS

The analytical process was ready to gather data for each pilot leader on the first two levels (pilot and topic research) and, on the third level we were ready to look for the possible correlations, dynamical development, mapping and eventually measuring relationships of various parameters.

The pilots being built for really various topics, technologies and learners, we have prepared several combinations of questions that showed "interesting" when going through selected pilots. These combinations were to be analyzed for individual pilot (e.g. do those who don`t like playing computer games and have good marks in STEM have also better achievement in knowledge test and did they improve more or less than those who do have lower marks?), for the whole topic (there were often more than one school in one topic) and for the whole dataset.

We could also look for the other possible splits – gender, age, habitat, attitude to school, etc.…

Our dataset showed really promising and we were to find the best suited way of data representation and analysis for the best possible interpretation.

The following describes the achievements and experiences of the analytical process.

3.1 System for gathering information and analytics

From the perspective of the current creation of information systems, it is a system with a standard division into several parts:

Part 1: Front-end

The Front-end part made in programming language PHP version 7 using HTML, CSS and JavaScript functionality. This approach allows for easy portability and sufficient functionality in relation to analytical functions. It has also been shown that, in spite of the initial worries about a potentially lower performance for analytical calculations, this solution, despite the large volume of data, is fully functional. Another important aspect is the possibility of integrating communication with other software modules and external software such as the LMS system. In this case, the system is equipped with a web services interface that is industrial standard.

Part 2: Back-end

The Back-end part was made in C# and running as a service that performs calculations that simplify and speed up the display. A typical example is the pre-counting of correlation tables. Another case is the processing of scanned questionnaires (OCR functions).

Part 3: Database

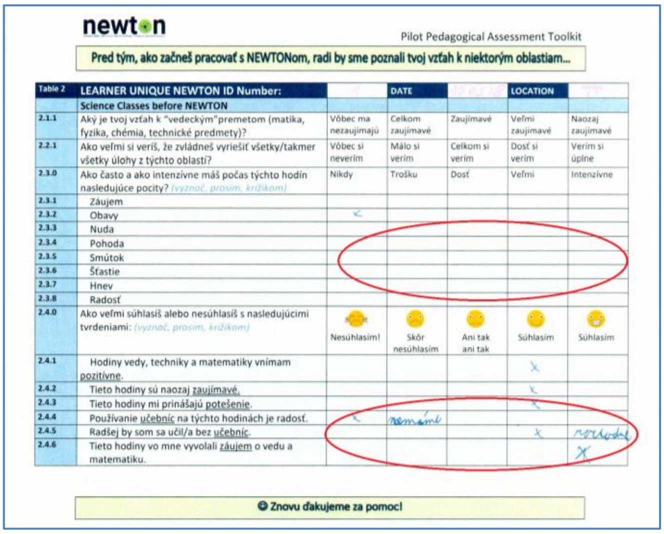
As a database server, a MySQL database has been chosen, which is otherwise sufficient for performance purposes.

The last, internal part is then the part of the system that is intended for the mathematical / analytical model. As the data contained in the database turned out to be quite suitable for modelling. Data can be divided into input (before the course) and output (after the course). Subsequently, based on the system theory, it is possible to identify the model of relational relation between input-output. In this case, the variability of the various models and the interpretation of the data, including, for example, extrapolation for cases not contained in the database, are extremely inspiring.

3.2 Data gathering

The original aim of the process of acquiring data from students and teachers was to enter this data electronically. Based on the practical experience of classroom equipment in different countries of the European Union, it has been shown that a more efficient way will be to prepare a questionnaire in printed form and then process it. Questionnaires were standardized using graphics to simplify and use by younger students, while translating the language of places where individual pilot courses took place until the process was done as planned and without any problems.

There was a presumption that the questionnaires obtained would be digitized or converted to standard XML forms using OCR and then automatically uploaded to databases. However, manual filling in of these questionnaires proved to be a critical moment, when some students, for example, checked more options and other non-standard situations that could not be electronically handled for this reason, it was necessary to proceed with a partially better processing of the data contained in the form.


newton

Pilot Pedagogical Assessment Toolkit

Pred tým, ako začneš pracovať s NEWTONom, radi by sme poznali tvoj vzťah k niektorým oblastiam...

Table 2	LEARNER UNIQUE NEWTON ID Number:		DATE		LOCATION	
	Science Classes before NEWTON					
2.1.1	Aký je tvoj vzťah k "vedeckým"premetom (matika, fyzika, chémia, technické predmety)?	Vôbec ma nezaujímajú	Celkom zaujímavé	Zaujímavé	Veľmi zaujímavé	Naozaj zaujímavé
2.2.1	Ako veľmi si veríš, že zvládneš vyriešiť všetky/takmer všetky úlohy z týchto oblastí?	Vôbec si neverím	Málo si verím	Celkom si verím	Dosť si verím	Verím si úplne
2.3.0	Ako často a ako intenzívne máš počas týchto hodín nasledujúce pocity? (vyznač, prosím, krížikom)	Nikdy	Trošku	Dosť	Veľmi	Intenzívne
2.3.1	Záujem					
2.3.2	Obavy					
2.3.3	Nuda					
2.3.4	Pohoda					
2.3.5	Smútok					
2.3.6	Šťastie					
2.3.7	Hnev					
2.3.8	Radosť					
2.4.0	Ako veľmi súhlasíš alebo nesúhlasíš s nasledujúcimi tvrdeniami: (vyznač, prosím, krížikom)	Nesúhlasím!	Skôr nesúhlasím	Ani tak ani tak	Súhlasím	Súhlasím
2.4.1	Hodiny vedy, techniky a matematiky vnímam pozitívne.					
2.4.2	Tieto hodiny sú naozaj zaujímavé.					
2.4.3	Tieto hodiny mi prinášajú potešenie.					
2.4.4	Používanie učebníc na týchto hodinách je radosť.					
2.4.5	Radšej by som sa učil/a bez učebníc.					
2.4.6	Tieto hodiny vo mne vyvolali záujem o vedu a matematiku.					

Znovu ďakujeme za pomoc!

Figure 1. Example of badly filled questionnaire.

In some cases, the data thus obtained can be corrected by consideration. In other cases, however, it is necessary to adjust the evaluation system in this lower data quality. The reason for this is that, for example, it is not possible to recognize in quantitative terms what the student's actual answer is. Therefore, the system was retrospectively modified to accept more options in some cases, and then statistically processed the more likely (chosen by multiple students of the course), or in other cases, calculated the average and adjusted the response accordingly.

As part of the questionnaire, there were not only checkboxes but, in a few cases, also the ability to write a short commentary on the course. For the evaluation in this section, the option of analysing the present keywords was chosen due to the fact that this filling took place in different languages and also these parts had to be translated into a common language that is English.

As a qualitative shift, the connection and export possibilities with the Newtelp LMS system proved to be better when the input was qualitatively better than when the questionnaire was digitized.

This is to alert researchers working with this kind of resources – as remedy we see here the strong emotional engagement in the subject (if not in the research itself). As example we mention here the Elecrophysics course (2 secondary schools, one vocational school from the small town, one big school for integration of children with special educational needs from the capital) where each of 33 learners furnished 2 valid questionnaires, knowledge tests and verbal suggestions and interviews. They happened to be emotionally engaged and appreciating their opportunity to contribute to the development of new education. The emotional engagement and the “right story” around the experiment can eventually attract much better and accurate work of any learner (in our sample there were socially excluded kids living in small villages, autistic adolescents, physically disabled children, posh white upper class adolescents and children, children from Hungarian and Roma minorities, aged from 15 to 19 years).

The original state of hand-filled questionnaires, combined with the insufficient emotional interest and concentration when completing them, did diminish the quality of the whole dataset by insignificant measure. It` s really not possible to reach the perfect match between input (original sheets) and output (dataset) when scanning the incomplete or poorly completed questionnaires or tests. But, the only other possibility being the manual upload of the data – for such an amount of data quite dangerous, the human data entry error making significantly more damages than machine data entry, no importance how sophisticated this manual entry will be [7].

When we compared the “unemotional” to “emotional” responses, there is a very important difference in the ratio of valid/invalid items. There are studies made on the importance of an attitude to the subject (e.g. [8] and [9]) but we are missing studies or research working with the importance of emotional engagement (or story) for research inputs of a high quality. To reach the highest quality of inputs possible, we suggest to find the right story (the role of local researcher emphasized) and focus the emotional side before the administration to gather the best possible cognitive outputs later.

3.3 Analytics

After the difficult transcription of several questionnaires we had to find the right way to represent the data. In the questionnaires, all answers were represented by the Likert scale 1-5. Unfortunately, the questionnaire was designed with the “serious” researcher on mind but, as it happened, these were only children without the strong attachment to the subject – the middle (3), neutral choice was selected far too often not to influence the whole result. This deforming bias could at least visually disappear when normalizing the data.

The scale 1-5 was great for filling the form but was totally non-illustrating. There were basically the questions of two types:

1. going from the worst (1) to the best (5): “How often do you feel enthusiastic at school: never – rarely – sometimes – often – always

2. the opposite – from the best (1) to the worst (5): “How often do you feel bored at school: never – rarely – sometimes – often – always”

When keeping the original scale, we don’t see any emotional or qualitative charge in the score – it does not bear any “significance”, it`s just lower or higher, and yet this varies too as we see the same values having the opposite meaning.

To include the true qualitative dimension in the data analysis, we decided to normalize the scoring to the other scale, from -2 to 2, going through -1, 0 and 1. We had also inversed the scoring for the second type of questions, to keep the most negative value on the negative side (-2) and the most positive value on the positive side (2). This brought us the “emotionally coloured image” what we considered indispensable for the research purpose. Attitudes, moods and opinions are emotionally founded, influenced and conditioned therefore, on our opinion, can`t be correctly analysed and further interpreted without this nuance.

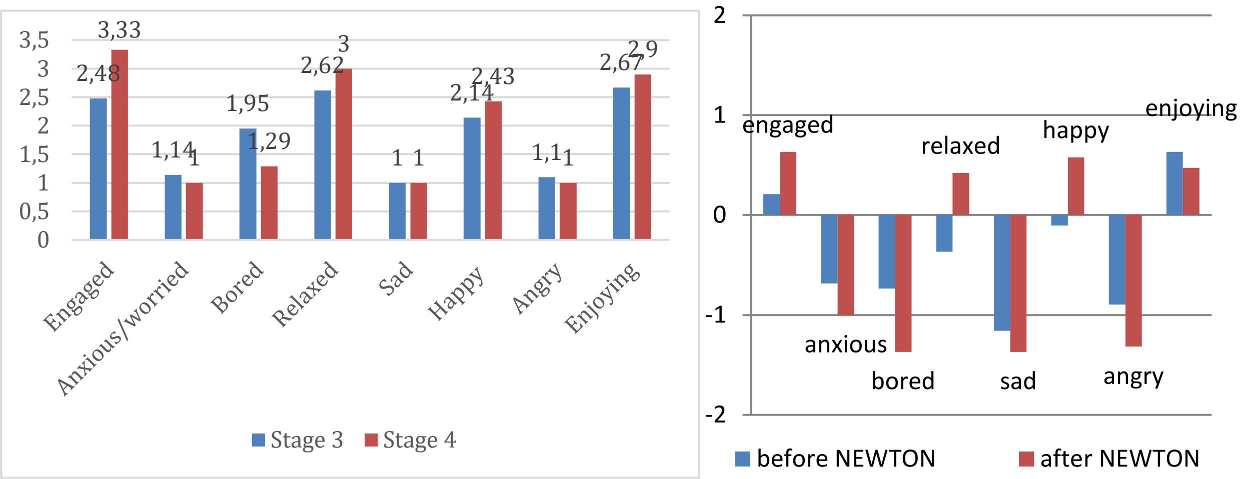


Figure 2. Graph with original and normalised results.

The limited extent of this paper does not allow us to show more points to be redressed in this kind of research – in the 3.3.2 we will describe more issues with the academical, non-data analytics-oriented researchers. The scale change was only one example in the possibility to gather more from the collected data. The raw dataset is more like music sheet than like music. It can be cleaned and normalized using the legal means, sorted or split or grouped and regrouped to show its riches. Other ways we`ll only apply mechanical methods, aka excel functions and the result will only bring the superficial conclusions (averages, correlations and t-test, none of them guaranteeing itself any sense and profit).

3.3.1 Technical advantages for used system approach

As it turned out during the project, the chosen technical solution brings many advantages. One of the most important are the very wide possibilities of filtering and formatting data before processing it for analytics. This is due to the use of very complex SQL queries that allow data to be interpreted over the entire database structure. Therefore, it is easier to separate the data preparation time before it is processed, which in principle is no longer possible in Microsoft Excel office applications. One example we can demonstrate follow.

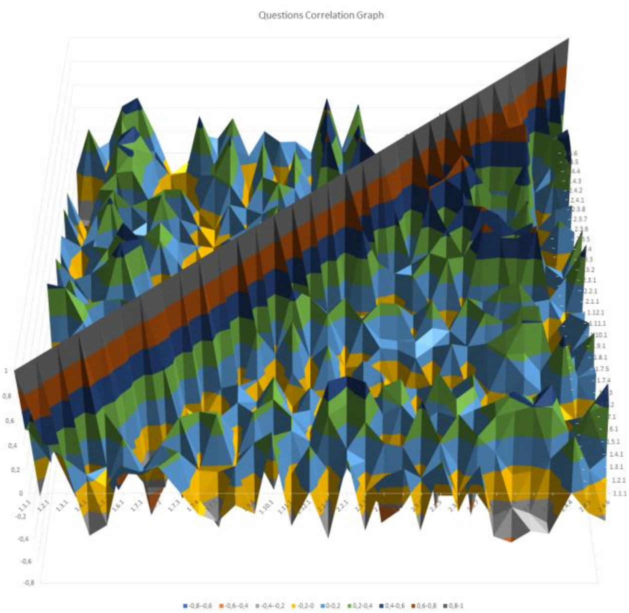


Figure 3. Correlation graph - Z axis is correlation coef. For questions results (X and Y axes). Higher peak means better correlation.

One of the advantages of a good database is the possibility of qualitative and quantitative work with the graphical output. To take advantage of this option, we have calculated a correlation coefficient according to Pearson for all questions in the system (but of course, a different approach, including T-test, etc.). These results can then be displayed in a 3D graph showing the correlation between the questions in a clear and clear way - a higher value towards 1 means a greater correlation. Obviously, all the correlation inputs were standardized so that the individual coefficients were adequate, for example, positive and negative questions were taken into account and they were standardized in one direction. The 3D graphs created in this way are able to calculate the analytical tool for different selections into the questionnaire and thus further work with correlations.

3.3.2 Practical experience for data analytics – social aspects

One of the surprising aspects that should be taken into account is how experts are able to evaluate the information they receive. Practical experience suggests that, based on the education of individual experts, the ability to analyse course results using "database" systems (means web environment) is very different. For example, it turned out that it is imperative that data be accessed from Excel tables because in the background it is an expert's education that is often based on using Excel as an analytical tool, however technically limiting it seems.

A similar problem also applies to the use of more advanced types of mathematical analytical apparatus for evaluating results. In fact, the university background often does not count on the use of such a more complex apparatus, however it may be essential for the assessment of the overall benefits.

As mentioned above, we had heavy difficulties to explain the usefulness of the system data analysis to the academical experts. There was no chance to show the possibility of more general approach, various data types and the work becoming possible thanks to this. Their own limits were even stronger than those of Excel itself – not being experienced in the more sophisticated use of Excel functions, it was visibly hard to see the "meaning" of the scores, points or numbers resulting.

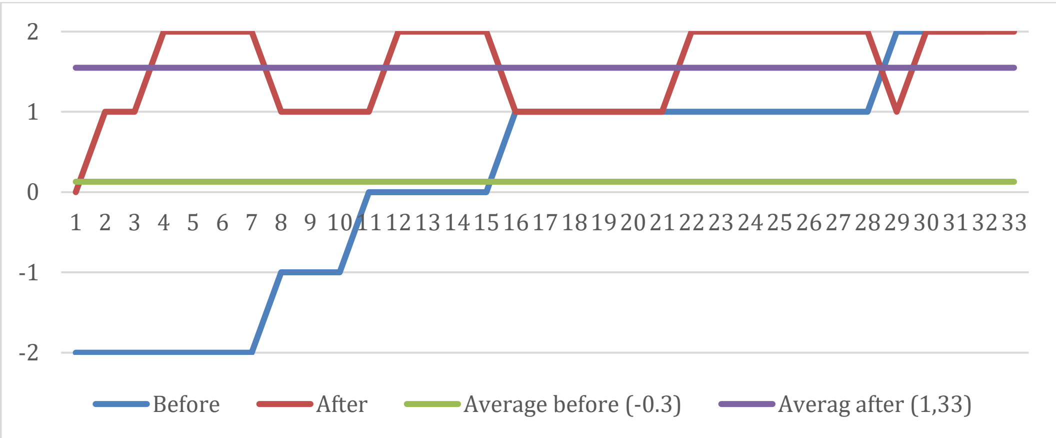


Figure 4. Example graph from evaluated pilot course.

The apparatus allowing researchers to work with any grouping or crossing or sorting or tree-developing remained unfortunately almost unused. We have made some attempts to show possible research/finding approach for one pilot but, lacking technical background on the pilot leaders` side, no general application followed.

3.3.3 Existing analytical approach

We can affirm there was a huge amount of analytical work relayed in the frame of the NEWTON project. Each local researcher had made her/his analysis following the requested groupings, using requested methods (average, t-test) on these groupings and making conclusions for each part of questionnaire divided by given sub-groups…

Some pilots provided interesting (statistically significant, approved by the *t-test*) results for almost all sub-groups, some pilots just for some sub-groups and, there were also pilots with no statistically significant results for any grouping. Then, there are other groupings, made by primary, secondary and tertiary education, each apart.

We see many partial results for specified datasets with no relation to any other group. It is not a dataset or Excel`s fault. The large set of partial results is not an issue itself. We see that almost every pilot in almost every school showed that children started to work with at least a little bit more interest than before.

But we don`t know if this improvement is connected to their age, school, technology, topic or anything else.

The size of the dataset was “screaming” for the full, complex analysis. The whole dataset (as mentioned already, more than 1100 participants to 32 pilots from 8 countries, 2 large questionnaires with more than 100 questions each (plus 3-4 small questionnaires for multi-lessons pilots, after each lesson), 2 knowledge tests (plus 3-4 small tests for multi-lessons pilots) – it was a beautiful large dataset with the treasure of collected data.

3.3.4 Other analytical approach / cross analysis

When analysing the similar dataset, the most interesting part comes after the whole set cleaned, ranged and, eventually normalized – ready to be screwed to reach important results, confirming, supporting or refuting the expectations, or, more scientifically, hypothesis.

In this case, we can show that as good the partially (by pilot) made analysis was, as insufficient they were for showing the impact of the whole project, utility of the platform or the pilots, attitude of learners or teachers.

When grinding the whole dataset in the system, we can then apply any filter we wish. We can go first over the whole dataset to look for correspondences, coincidences, identities, then to test originally given combinations of factors/parameters/questions by, again, any factor or other combination of factors across all pilots, selected pilots, all topics, selected topics, etc. And, the very best, it does not take any considerable amount of time. For the data analytics is used to work with dataset, by definition – and, in this dataset’s size, the chewing time bordering zero, we could try literally almost any combination in no time.

After having seen these calculations, it was obvious to look for the connections and possible correlations, or, better, just to choose from amongst already calculated results those showing the most relevant results.

4 CONCLUSIONS.

The aim of this paper was to show some possible issues of the educational data analysis and their possible solution. We think that for obtaining the best dataset possible, it`s obvious not only to well prepare the basic analytical tools and sample, but also not to forget to determine the importance (or possible impact) of sufficient emotional engagement (motivation) and the way of administration, be it paper, screen or speech. On our experience with NEWTON, we can say that each approach brings its strengths and con`s.

The initial speech of the national (at least!) project coordinator gives the allure to the pupils= work, enough even for teen-agers; hand-filled forms socialise learners (at the point they have discussed their school lessons much more openly (after having worked with our pilots)). On the other hand, machine data entry can help eliminate any emotional engagement (sometimes needed too), can accelerate the analysis and eliminate any social debate.

On our experience, the three dimensions – knowledge, attitude, technology – are emotionally bonded and therefore can`t be researched without considering this point, across all other factors and relevant groupings.

Where emotions are in the focus, attention gets sharper and dataset becomes potentially useful and, for those who see, beautiful [11]. From all its factettes, across all possible groups or pilots. Vivat structures, correlations and/with/by emotions!

ACKNOWLEDGEMENTS

Optional statement to thank other contributors, assistance, or financial support.

REFERENCES

[1] http://www.newtonproject.eu/

[2] de Reuver, Mark & Sørensen, Carsten & Basole, Rahul. “The digital platform: a research agenda”. Journal of Information Technology. 2017;10.1057/s41265-016-0033-3.

[3] Kärkkäinen, K. and S. Vincent-Lancrin, "Sparking innovation in STEM education with technology and collaboration: A case study of the HP Catalyst Initiative", OECD Education Working Papers, 2013

[4] OECD, Measuring Innovation in Education: A New Perspective, OECD Publishing, Paris, 2014. http://dx.doi.org/10.1787/9789264215696-en.

[5] Lin M, Chen H, Liu K. “A Study of the Effects of Digital Learning on Learning Motivation and Learning Outcome“. Eurasia Journal of Mathematics, Science and Technology Education. 2017;13(7):3553-3564. doi:10.12973/eurasia.2017.00744a.

[6] ALyu D, Wang B. “Effects of the Application of Computer Network Technology to Guided Discovery Teaching on Learning Achievement and Outcome”. Eurasia Journal of Mathematics, Science and Technology Education. 2018;14(7):3269-3276. doi:10.29333/ejmste/91249.

[7] Barchard, Kimberly & Pace, Larry. Preventing human error: The impact of data entry methods on data accuracy and statistical results. Computers in Human Behavior. 27. 1834-1839. 2011. 10.1016/j.chb.2011.04.004.

[8] Bal-Taştan S, Davoudi S M M, Masalimova AR, et al. “The Impacts of Teacher’s Efficacy and Motivation on Student’s Academic Achievement in Science Education among Secondary and High School Students“. Eurasia Journal of Mathematics, Science and Technology Education. 2018;14(6):2353-2366. doi:10.29333/ejmste/89579.

[9] Kwon, H. (2016). Effect of Middle School Students’ Motivation to Learn Technology on Their Attitudes toward Engineering. Eurasia Journal of Mathematics, Science and Technology Education, 12(9), 2281-2294.

[10] Llbao, N., Sagun, J., Tamangan, E., Pattalitan, A., Dupa, M., & Bautista, R. (2016). Science learning motivation as correlate of students’ academic performances. Journal of Technology and Science Education, 6(3), 209-218.

[11] Goleman, D., Focus: The hidden driver of excellence. New York, NY, US: HarperCollins Publishers. 2013.