Supporting the Acquisition of Scientific Skills by the Use of Learning Analytics

Daniel J. Salas[1(✉)], Silvia Baldiris[2,4], Ramón Fabregat[3], and Sabine Graf[4]

[1] Socrates Group, University of Córdoba, Montería, Colombia
danielsalas@correo.unicordoba.edu.co
[2] Direction of Research, Innovation and Social Projection, Fundación Universitaria Tecnológico Comfenalco, Cartagena, Colombia
sbaldiris@tecnologicocomfenalco.edu.co
[3] Institute of Informatics and Applications, University of Girona, Girona, Spain
ramon.fabregat@udg.edu
[4] School of Computing and Information Systems, Athabasca University, Edmonton, Canada
sabineg@athabascau.ca

Abstract. Beginning researchers in general face various difficulties when initiating a process of scientific research due to the unavailability of proper tutoring or the minimum knowledge about research methodology and, this impacts the reliability of the process, the time and the results of the research in question. The purpose of this work is to support the acquisition of scientific skills by offering to beginning researchers learning analytics in each and every one of the phases and stages of the investigative process based on the actions and interactions that teachers/supervisors, experts and researchers make during this investigative process. Therefor, it is presented, as a detailed case study, the skill of formulating research questions by defining the process that was used, including the actors, the measurements, and the indicators, the formative process and the interactions managed with the Binnproject software. Finally the K-means algorithm is used in analyzing students' behavior and creating clusters according to their performance during the process of formulating scientific questions, this way supporting the process of determining strategies able to strengthen scientific competences for both students and the teaching practice.

Keywords: Learning analytics · Scientific skills · Research questions · K-means algorithm

1 Introduction

Generally, the main difficulties that beginning researchers face within investigative processes have to do with the little availability of tutors to guide the young researchers, the lack of conceptual clarity regarding the research methods and techniques, difficulties reviewing the scientific literature, particularly related with the systematization of the gathered information, as well as problems with the management of bibliographical sources. Added up to all these, also the feeble ability to set the research design, both

D.K.W. Chiu et al. (Eds.): ICWL 2016, LNCS 10013, pp. 281–293, 2016.
DOI: 10.1007/978-3-319-47440-3_32

concerning qualitative and quantitative approaches, intensified issues regarding the analysis and interpretation of the data gathered during the research and the evaluation scenarios, and regarding the conditions and the manner in which these should develop considering the types of research and context situations [1].

The aforementioned issues require new monitoring approaches and innovative strategies pertaining to measurement systems that incorporate elements enabling regular self-reflections about the training process, along with efficient tutoring that includes early alerts at all times for students, as well as facilitating the mechanisms and the instruments that help reach all the competences and skills that the training process requires. In this sense, the experience relating with implementing learning analytics, such as the work that University of Purdue developed through the Signals platform [2], University of Athabasca's experience with the AAT tool [3], among others, shows that learning analytics are useful conceptual and technological tools given the capabilities they have to support effectively all training processes.

This article introduces a proposal for learning analytics supporting the acquisition of scientific skills, which was achieved through analyzing and reflecting about the research process.

A case study is presented, related to the skill of formulating research questions, establishing measurements and indicators that reflect the potential advancements of beginning researchers, implemented through a training process during the "Formulating scientific questions" module. This was developed within a group of 30 students, having to do with the conceptualization of scientific inquiries and aiming to attain an appropriation of the theme through the exercise of creating a classroom project with the Binnproject software. This software holds records of monitoring teacher/student interactions during the process of formulating research questions.

The article is structured in the following manner: the first section represents the introduction, the second a review of the state of the art regarding scientific skills and learning analytics, the third describe the scientific skills considered in this study, forth section presents our proposal of a learning analytics model to support the acquisition of scientific skills, a fifth section presents a specific case study to validate the model specifically considering the skill formulating research questions and finally, the conclusions and references are outlined.

2 State of the Art

2.1 Scientific Skills

At an international level, according to the guiding vision in training US youngsters, the abilities of the XXI century [4] have been determined to be, mainly, critical thinking, scientific alphabetization [5], innovation and collaboration, among others.

Europe and Latin America have been implementing and pushing the TUNING project [6, 7], which defined the scientific skills needed by undergraduate students in cycles of higher education.

Australia's experience [8] is outstanding as well, considering the Ministry of Education determined strategies with encouraging results in strengthening scientific skills.

On the other hand, as a multilateral organization, UNESCO [9] pointed out that scientific competences are a significant tool in solving social issues, therefor driving scientific alphabetization in the world.

From another angle, OCDE [10] considers that citizens are more productive and have increased problem solving capabilities when they benefit from scientific competences, and in this sense, countries members of this organization have been pushing PISA experiments geared for teenagers between the ages of 15 and 18.

One of the biggest challenges associated with the acquisition of scientific skills by youngsters of different educational institutions are the measurement tools [11], as well as the monitoring tools used in effectively assessing students' advancements [12]. With regard to this, learning analytics can be useful to observe student's behavior and alert them about their performance during the training process, aiming to succeed in both conceptual and practical understanding about scientific competences and skills.

2.2 Learning Analytics

Learning Analytics, as an emergent research area, has its bases in the following concepts: Big Data [13], Business Intelligent (BI) and Data Mining (DM). This latter concept is related to the extended term Educational Data Mining (EDM) [14, 15].

The emergence of the Learning Analytics concept has been motivated by the massive use of Virtual Learning Environments (VLE), Adaptive Hypermedia Systems (SHA) and Personal Learning Environments (PLE), as well as by advancements in computation and by ubiquitous learning, generating a large quantity of data and information as a consequence of the interactions between students, teachers, staff and the several stakeholders interested in these types of systems.

There are various definitions, widely known, about Learning Analytics, for example those by Siemens [16, 17], Elias [18], Duval [19] and Shum [20]. A commonly accepted definition is the one reached during the Learning Analytics and Knowledge (LAK) conference, characterizing this term as "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs".

Nevertheless, learning analytics as an emerging research area has generated different approaches and lines of work from a social learning [21] perspective, that was also set out and developed by Shum [22]. These studies define five categories of analytics, divided into two groups. The first shaped by social networks analysis and discourse analysis. The second shaped by content analysis, motivation and context analysis.

The acquisition of scientific skills can be beneficiated by a social learning process [23]. Supporting this process by offering the adequate technology to collaborate among the actors in the learning process as well as by supporting the process means learning analytics offered to student can help the students to advance in the acquisition of scientific skills in the following ways:

1. Monitoring the process of advancing in the acquisition of scientific skills
2. Predicting beginning researchers who are at risk of not achieving scientific skills.

3. Generation of early alerts that support the implementation of teaching strategies to help beginning researchers when necessary during the acquisition of scientific skills.

The identification of skills required in different areas and to empower teachers and students to use learning analytics to support their own achievement has been defined by Rebecca Ferguson as part of the future of learning analytics [24].

3 Definition of Scientific Skills

Students' skills can be inferred and monitored by the use of Learning analytics. The first step of this study was to define a set of abilities able to describe the advancements of beginning researchers in acquiring scientific skill. After studying and analyzing scientific skills frameworks [9, 10, 25] the following set of skills were defined:

A1 – Quality of the inquiry process and the main researchers expertize

- S1 – The ability to observe, recognize and detect particularities in phenomena, problematic situations, contexts etc.
- S2 – The ability to raise the research's main problem starting from a set of observations, data and theoretical references
- S3 – The ability to formulate research questions

A2 – Concerning the quality of the literature revision and the conceptual and practical appropriation by training researchers

- S4 – The ability to methodologically structure a revision of the specialized literature process, starting from specific research questions
- S5 – The ability to research specialized literature both in national and international reputable sources
- S6 – The ability to judge the quality of specific literature considering well defined criteria
- S7 – The ability to synthetize data of specific literature by categorizing
- S8 – The ability to report the findings from literature revision

A3 – Inherent to students' capacity to set objectives, justify the research, as well as set the approach and the scope of the research

- S8 – The ability to correctly determine the objectives of the research
- S9 – The ability to justify the research both from a theoretical and practical point of view
- S10 – The ability to set the approach and the scope of the research

A4 – Related to the training researchers' ability to set a quality research design

- S11 – The ability to set appropriate research designs according to the objectives, the approach and the scope of the research

A5 – Related to quality data collection and analysis by junior researchers

- S12 – The ability to use appropriate data collection techniques considering context variables that support getting reliable data
- S13 – The ability to create data collecting instruments
- S14 – The ability to identify the most adequate techniques in analyzing data in a research
- S15 – The ability to handle technological tools in order to analyze data in a research

A6 – Related to the quality of data interpretation and the scientific reports written by junior researchers

- S16 – The ability to interpret results and adequately represent them
- S17 – The ability to adequately report findings and conclusions of the research

Based on the defined skills, a process to develop measurements and indicators able to assess the learning advancements and continuous self-reflection of beginning researchers was carried out and it is introduced in the next section.

4 Learning Analytic Model

The interest of this study is to define a process based on learning analytics to support the students' acquisition of scientific skills as well as to support decisions of the teachers orienting students in the learning process. This process consist in the following steps:

1. To select the skill to be supported
2. To define the actors who interact in the process of scientific competences acquisition.
3. Determining measurements and indicators considering the learning/teaching process.

In this section we center our attention on the skill called: "The ability to formulate research questions" to describe the whole process in the following paragraphs.

4.1 The Skill Selected

The skill under examination is related to the process of formulating questions during a research

- S3 – The ability to formulate research questions

The acquisition of this skill permits the student to be able for the identification, analysis, characterization and formulation of research questions.

4.2 Actors During the Process of Scientific Skills Acquisition

The main actors involved in the process of scientific skills acquisition are:

- Experts in research methodology, who, based on their experience, support the definition of the scientific skills, as well as the establishment of measurements and indicators.
- Students or beginning researchers interested in acquiring scientific skills.
- Teachers that could play the role of director or co-director who advice a research.
- External experts freely participate during the learning and teaching process.

4.3 Description of the Training

For the particular case of "The ability to formulate research questions", the following training was defined. The training consist of 15 days, working 2 h every day using Moodle platform where the course called "Formulating research questions" was created. A teacher of the system engineering program, experienced in developing scientific research projects, is in charged to offer the training. The training proposes the development of activities related to scientific questions, particularly creating the questions considering their characteristics, their relevance and their pertinence in the context of a research.

Additional to some lectures, the training includes discussion forums, addressing aspects one has to take into consideration when formulating research questions, as well as an assessment to evaluate the topics understanding.

The training also includes an activity to identify into scientific literature the research questions which were defined and addressed in each particular paper. Selected papers that comes from areas related with the system engineering program, Ubiquitous computing, augmented reality and educational video games. In this way literature is familiar to the students.

4.4 Determining Indicators

In the context of this study, an indicator is defined, as "an individual or composite statistic that relates to a basic construct in education and is useful in a policy context" [26].

The indicators related to the skill of formulating research questions were defined by two experts in research methodology, they are shown into Table 1. The purpose of these indicators to describe the advancements or the appropriation of students or beginning researchers in formulating scientific questions.

Used measurements to generate the indicators are:

- Formulated Questions by a student (FQS)
- Formulated Questions by all the student (FQAS)
- Total Formulated Questions by a student (TFQ)
- High Quality' Formulated Questions (HQ-FQ)

Table 1. Measurements and indicators

No	Indicator	Used measurements to define the indicators	Description	The reach of the skill
1	%P = FQS/FQAS	FQS	Indicator of the participation of the student in formulating scientific questions during the learning process	The student makes scientific questions
2	QFQ = HQ-FQ/FQS	FQS RFQ	Indicator of Formulating Quality. All questions specified by the students are assessed by teachers and in this way they could be (High Quality Question (HQ-FQ), Medium Quality Question (MQ-FQ) and Low Quality Question (LQ-FQ))	The student makes scientific questions
3	%CFQ = CFQ/TFQ %AFQ = AFQ/TFQ %RFQ = RFQ/TFQ	CFQ RFQ RFQ	Indicator of the participation of the teacher during the learning and teaching process. It indicates teacher implications in analyzes and assess the quality of the questions the student formulates suggesting improvements	The teacher actively participates during the learning and teaching process revising the process of creating these questions. As mentioned before, Teachers that could play the role of director, co-director or external experts who advice a research
4	%CFQ = CFQ/TFQ %AFQ = AFQ/TFQ %RFQ = RFQ/TFQ	RFQ	Indicator of the participation of experts during the learning and teaching process. It indicates teacher implications in analyzes and assess	The expert is an external actor valuing students' contributions and making judgments and assessments

(*Continued*)

Table 1. (*Continued*)

No	Indicator	Used measurements to define the indicators	Description	The reach of the skill
			the quality of the questions the student formulates suggesting improvements	
5	%Q-I = I-RFQ/RFQ	I-RFQ RFQ	Is related to the student processing the considerations of the director, co-director and expert. The adjustments made by the student are being furthermore revised and valued till the final question is specified	Progress in creating research questions is being observed, as a consequence of the judgments and assessments made by the teacher and internalized by the students

- Reviewed Formulated Questions by teachers or experts (RFQ)
- Commented Formulated Questions by teachers or experts (CFQ)
- Assessed Formulated Questions by teachers and experts (AFQ)
- Action Recommendation on the Formulated Questions by teachers or experts (RFQ)
- Quality of the Interaction (Q-I)
- Improved of Reviewed Formulated Questions (I-RFQ)

Described indicators are aligned with the quality indicators for learning analytics defined by Sheffel [27]. In particular with the learning support indicators and learning measures and outputs.

5 Validation Scenario

With the purpose to validate the Learning Analytic Model the following process was defined:

1. The development of BinnProject module to manage research projects over Moodle.
2. Execute the designed training process with students or beginning researchers that wants to learn how to formulate research questions.
3. Results analysis including averages obtained by students in process of creating research questions and the application of K-Means algorithm to verify the students' behavior during the training formulate research questions.

5.1 BinnProject Module

Binnproject was implemented as a Moodle module to support students when they develop research projects. The work of the students over Binnproject is completely articulate with the training to formulate research questions described before.

Binnproject offers to students functionalities to collaborative editions of projects including the possibility to add reviews, comments or recommendations to be consider when they work in different elements of a project, for instance, research questions. Binnproject also manage notifications, the generation of the indicators, and several visualization including those related with the advance of the students in the learning and teaching process. Figure 1 shows and outline of the software functionality.

The functionality of collaborative edition of projects permit to define dynamic and personalized projects templates that could be adapted to specific needs of any institutions or research unit. From an edition mode the students are able to use the templates for projects creation.

Review functionality offers to teachers and experts generate comments, recommendations and to assess several elements of a project, among then, the formulated research questions.

Notifications functionality facilitates to inform to teachers, experts and students about events that happen in a project creation. For instance, students are notified when teachers review, commend or recommend actions to improve their projects.

Analytics functionality process the information related with the students, teachers and experts interaction to generate the defined indicators or other relevant data analysis.

Visualization functionality display graphically the information about the indicators and the students progress in the project development according with the planning previously stablished.

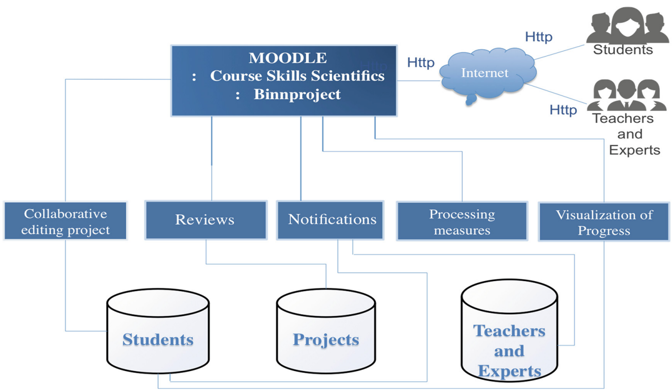

Fig. 1. Binnproject funtionality

5.2 Sample Involved in the Training

In the population study object 30 students participated a Training for beginning researchers. 21 males and 9 females, who were undergoing the eight semester of the system engineering program of the University of Córdoba in Colombia. The ages varied between 20 and 22 years old. They came from institutional beginner research centers (called “semilleros” in Spanish meaning “seedbeds”).

5.3 Results of the Training

This section describes the main results obtained by the students in the training for formulating research questions using Binnproject.

Actors Participation and Interaction. Students had an active participation in the discussion forums, 27 students participated with comments and supporting their thoughts about how to formulate research questions, main difficulties in the formulation and how to deal with they.

Regarding the activity of searching research questions into research papers, 30 students participated coming up with the following results: the average of scientific papers gathered by each student was of 10 papers, 90% of the students were able to identify the questions pertaining to the research in those respective papers and the other 10% of students had difficulties finding and describing the research question in their gathered papers.

Main research areas in the papers were software engineering, ubiquitous computation, augmented reality and educational video games.

Table 2. Commentaries and recommendations made by the teachers and the expert

Commentaries	Recommendations
1. Writing the questions in the infinitive tense	1. The question should be formulated using simple language
2. Revising the orthography	2. The question should be concise
3. The question is too general	3. The question should have answer options
4. Revising the verbs used to ask the question	4. Verify if the question is relevant
5. The question cannot by asked in a way that requires a yes/no answer	5. Determine the question of the research
6. Revising the coherence of the research question	6. The question should have a proper level of complexity
7. The question is ambiguous	7. Disaggregate the question
8. The question is obvious	8. Do not write questions as affirmations
9. It’s a question that generates an opinion and not a research question	9. Avoid excessive simplicity
10. The question is not clear enough	10. The question relates to multiple areas of study

The beginning researchers were asked that based on the papers they studied, gathered around the four major research areas: software engineering, ubiquitous computation, augmented reality and video games, to start formulating a research proposal by initiating a process of creating research questions using the BinnProjectt system.

Beginning researchers were organized in the following manner: eight (8) students were dedicated to formulate research questions in the software engineering area, six (6) students in the area of ubiquitous computation, eight (8) students in the augmented reality area and eight (8) students in the video game area.

During the process of formulating investigative questions in the BinnProjectt system, on average, this is what was generated: 10 commentaries and 2 recommendations by each revision of the director teacher, co-director teacher and expert, in charge of monitoring the questions being created. The main commentaries and recommendations types made by the teachers and the expert about the process of formulating research questions are described in Table 2.

Students' Scores. The main results achieved by students in the assessment graded by teachers and experts during the training are shown in Table 3. Table 3 shows no significant differences between the scores given to students by teachers and experts, however scores given by experts were slightly higher average than those give by teachers.

Table 3. Scores assigned by Teachers and experts to students

Teachers and experts	Attended Sample	Average Score assigned by Teachers and experts	Standard deviation	Variation coefficient	Min	Max
Co-director's	30	3,19	0,69894	21,9103%	2,0	5,0
Director's	30	3,05	0,844352	27,6837%	2,0	5,0
Expert's	30	3,24	0,906148	27,9675%	1,5	5,0
Total	90	3,16	0,815827	25,8173%	1,5	5,0

5.4 Weka Analysis

Based on information from the 30 students under study and the assessments made by the teachers and experts in each questions, the K-Means algorithm was applied in Weka in order to determine the behavior of the students on a group level, using four iterations, thirty observations and three variables, as well as defining two clusters, 0 and 1. The first cluster had the initial points [3,4,4] selected in a random manner and the second cluster with the initial points [4,3,3], also selected in the same way as the previous one (Table 4).

Related to the cluster 0 set up, it is equivalent to 57%, corresponding to 17 students, whereas within cluster 1, the 43% correspond to 13 students. Clusters group students according with their Low or High level of performances. It could help teachers to offer special guidance to students located in the Low Level.

Table 4. Results implementing the K-means algorithm in Weka

Attributes	Complete data (30)	Cluster 0 (17)	Cluster 1 (13)
Director's assessment	3,05	3,5882	2,3462
Co-director's assessment	3,19	3,6059	2,6462
Expert's assessment	3,24	3,764	2,5538

The results applying the K-Means algorithm indicate that evidently, the 13 junior researches grouped in cluster 1 obtained an assessment varying between 2,3 and 2,6, and that, on the other hand, it is observed that there is a group of 17 junior researchers that obtained a major average assessment, varying between 3,5 and 3,7. However, when observing all the collective data, that is, from all the 30 students, the assessment's relative average is between 3,05 and 3,24.

6 Conclusions

This work showed an outline of analytics supporting the acquisition of scientific skills by junior researchers. Equally, a complete process following a specific case study related to the skill of formulating research questions was presented, where the most significant results are related to conceptual appropriations by the junior researchers as a consequence of a training process. Another aspect achieved was a process of interaction between junior researchers and reviewers and experts through the BinnProject software, managing to collaboratively elaborate research questions within a research project. Through a process of regular and permanent revision, the advancements of the junior researchers were highlighted, obtaining as well feedback that facilitated reflection and analysis of the training process. Added up to this, an exercise of behavior analysis was conducted to the junior researchers, using the assessments made by the teachers and experts in regard to the research questions, adjusted to be examined in Weka using the K-means algorithm. There, it was observed that evidently, the behavior of a group of 17 junior researchers towards the assessments made to formulate research questions was relatively higher than the one in the second group of 13 researchers.

The final purpose is to generate new strategies to improve practices in the classroom.

References

1. Reeves, T.C.: Enhancing the Worth of Instructional Technology Research through "Design Experiments" and Other Development Research Strategies. Paper presented Annual Meeting of the American Educational Research Association, New Orleans (2000)
2. Pistilli, M.D., Arnold, K.E.: Course signals at Purdue: using learning analytics to increase student success. In: 2nd International Conference on Learning Analytics and Knowledge, May, pp. 2–5 (2012)

3. Graf, S., Ives, C., Rahman, N., Ferri, A.: AAT: a tool for accessing and analysing students' behaviour data in learning systems. In: 1st International Conference on Learning Analytics and Knowledge, pp. 174–179 (2011)
4. Burkhardt, G., Gunn, C., Dawson, M., Coughlin, E.: Literacy in the digital age. Br. J. Educ. Technol. **37**, 315 (2003)
5. Turiman, P., Omar, J., Daud, A.M., Osman, K.: Fostering the 21st century skills through scientific literacy and science process skills. Procedia Soc. Behav. Sci. **59**, 110–116 (2012)
6. González, J., Wagenaar, R.: Una introducción a Tuning Educational Structures in Europe. La contribución de las universidades al proceso de Bolonia. Bilbao Publicaciones la Univ. Deusto, p. 96 (2009)
7. Beneitone, P., Gonzalez, J., Siufi, R., Wagenaar, G.: Reflexiones y perspectivas de la educación superior en América Latina. Inf. Final. Tuning—América Lat., pp. 1–432 (2004)
8. Tobergte, D.R., Curtis, S.: National assessment program – science literacy year 6 report. J. Chem. Inf. Model. **53**(9), 1689–1699 (2013)
9. Gil Pérez, D., Macedo, B., Martínez, Torreglosa, J., Sigifredo, C., Valdés, C., Vilches, A.: Cómo promover el interés por la cultura científica (2005)
10. OECD: Pisa 2015 Draft Science Framework (2015)
11. Blikstein, P.: Using learning analytics to assess students' behavior in open-ended programming tasks. Learning, October, pp. 110–116 (2011)
12. Worsley, M., Blikstei, P.: What's an expert? Using learning analytics to identify emergent markers of expertise through automated speech, sentiment, and sketch analysis. Educ. Data Min. **2011**, 5 (2011)
13. Bollier, D.: The Promise and Peril of BIG DATA
14. Romero, C., Ventura, S.: Educational data mining: a survey from 1995 to 2005. Expert Syst. Appl. **33**(1), 135–146 (2007)
15. Siemens, G., Baker, R.S.J.: Learning Analytics and Educational Data Mining: Towards Communication and Collaboration (2010)
16. Siemens, G., Gasevic, D., Haythornthwaite, C., Dawson, S., Shum, S.B., Ferguson, R.: Open learning analytics: an integrated & modularized platform proposal to design, implement and evaluate an open platform to integrate heterogeneous learning analytics techniques (2011)
17. Siemens, G.: What are Learning Analytics? Pagine Web (2011)
18. Elias, T.: Learning analytics: definitions, processes and potential. Learning, p. 23 (2011)
19. Duval, E.: Learning Analytics and Educational Data Mining (2012)
20. Shum, S.B.: Learning analytics. J. Interact. Online Learn. Soc. Behav. **53**(4), 395 (2012)
21. Ferguso, R., Shum, S.B.: The Open University' s repository of research publications Social Learning Analytics: Five Approaches Conference Item Social Learning Analytics: Five Approaches (2012)
22. Shum, S.B., Ferguson, R.: Social learning analytics. KMi Technical Report (2011), vol. 15, June (2011), p. 2330616 (2012)
23. Ferguson, R., Shum, S.B.: Social Learning Analytics: Five Approaches, Open Univ.'s Repos. Res. Publ. (2012)
24. Fergunson, R.: Learning analytics: visions of the future. In: The "LASI Spain 2016" International Workshop is Organized by the University of Deusto (UD) with the Collaboration of SNOLA (Spanish Network of Learning Analytics) (2016)
25. NPA: National Postdoctoral Associate Core Competencies Self-Assessment Checklist
26. Shavelson, R., McDonnell, J.: What are educational indicators and indicator systems? Pract. Assessment Res. Eval. (1991)
27. Scheffel, M., Drachsler, H., Stoyanov, S., Spech, M.: Quality Indicators for Learning Analytics, vol. 17, pp. 124–140 (2014)