ISBN: 978-989-8533-81-4 © 2018

A MODEL FOR ENRICHING AUTOMATIC ASSESSMENT RESOURCES WITH FREE-TEXT ANNOTATIONS

Luigi Di Caro[1], Sergio Rabellino[1], Michele Fioravera[2] and Marina Marchisio[2]
[1]*Department of Computer Science, University of Turin, Turin (TO), Italy*
[2]*Department of Mathematics, University of Turin, Turin (TO), Italy*

ABSTRACT

This paper discusses a model for structuring resources for automatic assessment in scientific education by means of textual descriptions. This study aims to support instructors in extending teaching strategies and expanding formative assessment in virtual communities of practice. The strategy to achieve these goals involves the implementation of a system for adaptively providing the students with instructional activities based on Natural Language analysis and ontological knowledge bases. The model is presented, and the feasibility of its use is discussed taking into account the results collected from the experimentation with instructors.

KEYWORDS

Automatic Assessment, Descriptors, Learning Objectives, Mathematics Education, Natural Language, Communities

1. INTRODUCTION

The development of e-learning courses typically involves some basic steps, such as choosing the hosting environment, creating learning resources, and storing the resources in the courses. Resources description and metadata can help their management through repositories; however, learners could experience difficulties in choosing among similar resources: even if it is easy to find data, it could be a challenge to retrieve an e-learning material that satisfies their needs – especially for those who do not have the complete awareness of what is needed to know in advance. The role of instructors/editors of an online course is still crucial to help students. At the same time, even e-instructors are not able to provide learners with uniform help.

The present research is aimed at scientific disciplines, especially Mathematics, and considers resources for automatic assessment (Barana, et al, 2015). It investigates the possibility to implement a system that automatically groups resources by means of the elements indicating their intended competencies. Through this system, instructors will no longer exclusively design learning paths but they will also be able to focus on the specification of the single nodes of possible paths. The research is aimed at

- supporting instructors in extending teaching strategies by the retrieval of shared resources;
- fostering learners' formative assessment by automatic recommendation.

This study is conducted by proposing and experimenting a model for the description of the scope of a resource (as well as of prerequisites and expected results) with instructors participating to a virtual learning community. The model provides for Natural Language descriptions: those will be the information exploited by an automatic recommender system integrated to the learning management system hosting the community. The model features are presented, and the results collected from the experimentation of its use by the instructors are discussed.

2. STATE OF THE ART

The next subchapters present the two main aims that guide the system and model development, namely resources retrieval and formative assessment, and the context in which they are applied – the Communities – for enhancing competence-based education.

2.1 Resources Retrieval

A lot of work has already been done in order to facilitate the storage and retrieval of shared e-learning resources. Several examples of online repositories exist, such as Merlot (http://merlot.org), CeLeBraTe (http://celebrate.eun.org), Wisc-Online (http://wisc-online.com). In the field of competency-based education, European Projects have been activated to help instructors share learning objects in an open perspective (Ravotto, 2011). In the Italian context, operational strategies have been proposed (Trinchero, 2017) for

- defining observable indicators for the skills being trained;
- building learning paths;
- designing a teaching that is based on national guidelines goals but insists on a plurality of cognitive processes, educational activities and assessment, tests and certification of competence;
- defining the levels of competence to be included in the final certification forms.

To overcome the problems that can occur in e-learning systems where the learning path is determined solely by human experts, strategies based on Natural Language have been developed and experimented (Vincent et al., 2014). Ontologies-enhanced approaches showed how external domain knowledge can help organize a corpus of digital materials and construct comprehensive knowledge representations to facilitate online learning (Huang et al., 2015).

2.2 Formative Assessment

There is no universally shared definition of “competence” (Winterton et al., 2006). In this paper, competence means the ability to cope with a task, or a set of tasks, managing to activate and orchestrate internal, cognitive, affective and resources, and to use the external resources available in a coherent and fruitful way.

Formative assessment is the way learners use information from judgments about their work to improve their competence. Since the nineties, the concern about formative assessment has grown to cover one of the major issues in the educational research. Paul Black and Dylan Wiliam conceptualized formative assessment through the following five key strategies (Black & Wiliam, 2009; Wiliam & Thompson, 2007):

1. clarifying and sharing learning intentions and criteria for success;
2. engineering classroom discussions and learning tasks that elicit evidence of student understanding;
3. providing feedback that moves learners forward;
4. activating students as instructional resources;
5. activating students as the owners of their own learning.

Matching similarities among digital materials is crucial for building mapped data sets of entities and relationships across entities, which are useful for the enhancing of automatic formative assessment strategies. The possibility to perform such an operation with ontology-based methods has been discussed in (Di Caro et al., 2018a; Barana et al., 2018a; Di Caro et al., 2018b).

2.3 Community

The use of automatic assessment systems has been proven effective in Mathematics education (Barana & Marchisio, 2016) and formative assessment (Barana et al., 2018b) especially in the context of Communities (Ellerani & Parricchi, 2010; Demartini et al., 2013). A Community is a system where

- instructors (discipline experts) manage one or more courses dedicated to a group of learners;
- tutors (discipline and ICT experts) help instructors with the testing of innovative methodologies for teaching, the creation of digital materials, peer collaboration, the sharing of resources and best practices, the use of advanced tools integrated to the LMS that hosts the online courses;
- instructors and tutors agree upon a framework of competences expected to be achieved by the learners at the end of the learning process.

A national-wide example of Community is the “Problem Posing & Solving”. PP&S started as a project promoted by the Italian Ministry of Education, to ensure a growth in the information technology culture in teaching and learning in secondary school and building a "Problem posing & solving" culture (Brancaccio et al., 2014; Brancaccio et al., 2015). The project aims at developing an integrated training area that interconnects logic, Mathematics and Information Technology. The PP&S Community has adopted the following suite of technologies: an Advanced Computing Environment, an Automatic Assessment System and a web conference system integrated to the Virtual Learning Environment which hosts the Community.

The community of instructors who joined the SMART project is an example of Community in a European context (Brancaccio et al., 2016). It shares with PP&S the suite of tools and the methodology chosen for fostering Problem Posing and Problem Solving in different European countries. The Project offers a guided training, an accessible repository of free materials, and an open community to share practices.

3. THE MODEL

The features of an assignment reflect the educational strategy chosen by the instructor. Whatever the features, clarifying, sharing, and understanding learning intentions and criteria for success are fundamental requirements to activate learners' reasoning on their own objectives achievement and to stimulate the learner to undertake appropriate learning paths (Trinchero, 2014). It is proposed to associate digital materials to Natural Language descriptions which explicate the learning intentions and success criteria they were designed for. The model provides for the P-R-O triad of descriptors: Performance, Requirements, Objectives. The P-R-O triad elements are defined as follows, and meant to be included as metadata about the material:

- Performance (P): specific and essential description of what a student will be able to do as proof of having achieved the objective, which can be measured objectively.
- Requirements (R): declaration of objectives - considered as already achieved by the students - necessary and sufficient to provide the performance.
- Objectives (O): general description of the knowledge (to be acquired or recalled) and / or of the intellectual skills and abilities (to be developed) that students are required to demonstrate in order to be considered competent.

The three descriptors must be expressed in a student-centered perspective. Performance statement should explicate what the student will be able to do (that can be observed directly), the conditions under which the student will perform the task, the criteria for evaluating the student's performance, and optionally a degree of mastery needed. Requirements is the explicit declaration of previous knowledge and skills necessary and sufficient to achieve the Objectives of the material. Objectives, unlike Performance, is independent from the type of response of the material (since distinct types of questions can be exploited for testing an identical instructional objective). This statement should not simply describe a list of topics, that being too abstract and narrow, and not being restricted to lower-level cognitive skills.

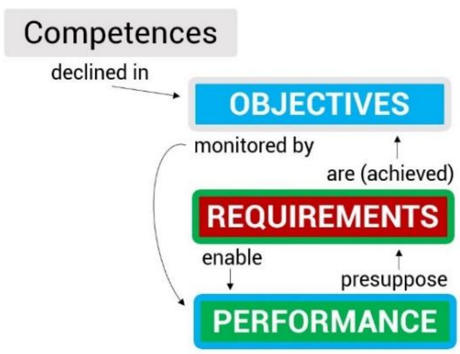

Figure 1. Relation between competences and the P-R-O descriptors

Figure 1 shows the P-R-O model relation with competences: competences can be declined in Objectives. The former can be monitored by a Performance, which is enabled by (presuppose) Requirements – which are intended as achieved Objectives.

Materials descriptors can express which the kind of performance required in terms of cognitive processes that need to be activated and the types of knowledge on which these processes operate. To this aim, the adoption of a taxonomic model is proposed. This is considered as the main reference affecting both instructors and learners: during the design phase, it is important to "space" in the definition of learning tasks; moreover, automatically subsuming cognitive processes and knowledge types implicit in a material is the key to adaptively advising students with variegated resources.

To detail the descriptors structure more deeply, this research considers Anderson & Krathwohl's taxonomy (Anderson et al., 2001). The authors proposed a classification of cognitive processes and knowledge types: 11 types of knowledge organized into 4 categories (Facts, Concepts, Procedures, Metacognition), and 19 basic processes organized into 6 categories (Remember, Understand, Apply, Analyze, Evaluate, Create) ordered by ascending cognitive complexity. Considering the authors' taxonomy, a material can be linked to a set of concepts couples referring to a two-dimensional matrix: the first dimension of the matrix represents the types of knowledge while the second dimension represents the cognitive processes involved. The connection between a material and a matrix element is established by identifying cognitive processes and knowledge type from its descriptors.

The use of Natural Language allows to overcome the limits implicit in the exclusive use of tagging, namely the inability to express complex data. The risk of the use of tagging is to achieve a semantic granularity that is either too synthetic or too high. In the domain of Natural Language it is possible to use a syntax from which to extract conceptual coherence: this allows to achieve both a greater freedom of expression by the designers and a minor effort for those who search for materials.

Inferring Anderson & Krathwohl taxonomy dimensions involves dealing with domain-specific verbs. The semantic disambiguation of a verb in terms of cognitive processes depends on the (disciplinary) element on which it is applied. The use of an ontological version of the taxonomy is proposed to be integrated with the OntoMath^{PRO} ontology, a bilingual ontology of mathematical knowledge, geared to be the hub for math knowledge in the Web of Data (Nevzorova et al., 2014). The developers share the sources with the Semantic Web community. This research proposes the further development and translation of OntoMath^{PRO} also in the Italian panorama.

4. EXPERIMENTATION

To validate and refine the model, quantitative and qualitative analyses on activities made by Community instructors is proposed. The main objective is to experiment the use of the model to equip the digital teaching materials with descriptors linked to the skills expected at the end of the educational path.

The experimentation requires a total estimated commitment of 6 hours within 4 weeks for each participant, and it develops through several activities within four main phases of inquiry:

- **Initial questionnaire**: inquiry about the methodology that instructors adopt to create materials.
- **Agreement activity**: evaluation of the consistency of the choice of descriptors with respect to the proposed model of sample questions. The activity follows a virtual meeting for the presentation of P-R-O descriptors model.
- **Workshop**: assessment of the descriptors – associated with questions – written by colleagues. The workshop follows the 'question design and submission' activity: each instructor autonomously elaborates 1 question starting from the specification of the descriptors. This activity is preceded by a 45-minute virtual meeting in which an example of design and implementation of a question (starting from the specification of the descriptors) is shown.
- **Final questionnaire**: inquiry about the activity carried out in the experimentation.

Questionnaires questions are open-answers or selections in Likert scale from 1 ("not at all") to 5 ("very").

5. RESULTS

Each of the next subchapters shows the results of the single phases of the experimentation. The subjects involved are secondary school teachers chosen among those who joined the PP&S project. To date, the Project involves 1,229 teachers (from all over Italy) of Mathematics, Computer Science, Physics, and other scientific disciplines, and 16,400 students. For this experiment, we chose a reduced set of (23) teachers, particularly active and proactive, dedicated to innovation and willing to experiment with new methodologies.

The experimentation did not provide for indemnity of any kind, it only issued a training certification by the University of Turin for each teacher involved.

5.1 Initial Questionnaire

The participants are not used to sharing content frequently: half of them shares less than 25% of the questions used with their students. At the same time, as Figure 2 shows, more than 50% of their questions are autonomously created. The other part of instructors' repository appears – on average – to be composed of questions taken from the Community (and possibly modified according to their needs).

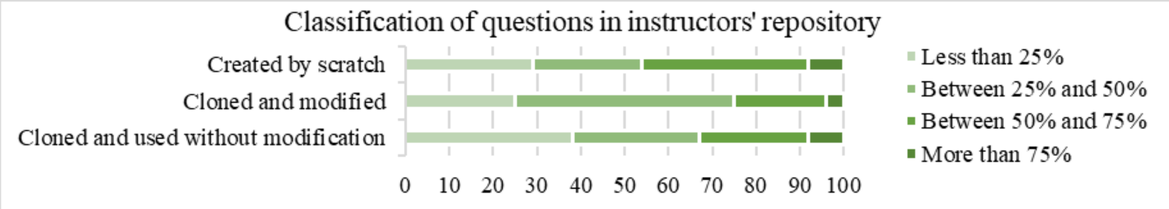

Figure 2. Percentage of questions created by scratch, cloned and modified, cloned and used without modification according to instructors' responses

The instructors' tendency to share a little percentage of their materials is accompanied by their low perception of the easiness of the search for questions in the Community: about 72% of them rates this aspect "enough" easy or less. They motivated their choices with open answers: among the features which contrast the easiness of the search engine they indicated the "heterogeneity of the quality of the questions", the frequent "poor specification of the topic of the questions", and the perceived "lack of immediacy in the research system". Some answers, like the latter, highly relate to the Automatic Assessment System features, while others – such as the former – deal with Community best practice: specifying topics and estimating the quality of shared materials should follow a common methodology in order to make it easy to retrieve. Topics can be part of the learning objectives description, which 80% of them specifies – mostly verbally – before submitting an assignment to students. Even among those who indicated that they do not specify learning objectives, the times of reflection about learning objectives during planning and realization of a question are rated "enough" or more. This tendency is similar when taking into account the time dedicated to reflecting about the way in which the question allows to measure the achievement of the objectives, and the requirements needed to answer the question. The rates slightly decrease when considering the phase of the use of questions with the students, in line with the presence of 20% of instructors who indicated that they do not specify learning objectives.

5.2 Agreement Activity

The survey on agreement of choice between different descriptors of proposed questions is proposed to assess instructors' understanding of the P-R-O model. The instructors are shown 4 questions, each of which has 4 different versions of each descriptor. For each of the 3 descriptors, each teacher is asked to "choose the sentence that appeals as the most appropriate for a given descriptor or suggest a different one"; the teacher can select one of the 4 choices or write his original proposal in the text area.

All instructors made the same choices, which matched an expert's opinion about which should be the most correct answers except for one case. This shows that the identification of distractors is an operation on which high degree of agreement can be achieved.

5.3 Workshop

Analyzing peer assessment rates and open answers enables to understand the following features:

- instructors' comprehension of the descriptors: Performance was the only one which someone misunderstood: it does not relate with the 'actual' solving path taken by the student. Instead, it describes what the student is required to do to communicate to the teacher the correct answer

(supposing that he already found it); for an example, see the review of teacher 2 to the Performance of the author A in Table 1;
- instructors' writing methods: the majority of them wrote essential statements (composed by one or two phrases), while others widely expressed descriptors using also lists of phrases;

The activity highlighted the understanding of the intent of the authors of the questions: several reviewers appreciated peers' work and proposed modifications to questions to completely align them with the descriptors.

Figure 3 shows an example of workshop submission, while Table 1 shows the related descriptors proposed by the author (A) and by the reviewers (1, 2, 3).

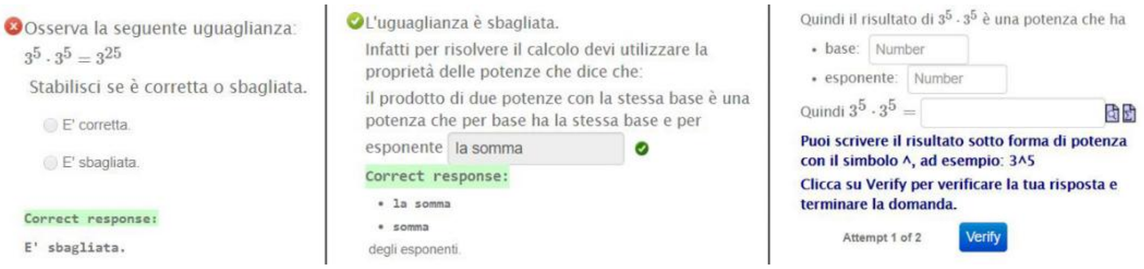

Figure 3. Question (composed of three parts) submitted by an instructor.

Table 1. Descriptors proposed by the author of the question (A) and by the reviewers (1, 2, 3)

	Performance	Requirements	Objectives
A	Assigned an already resolved expression (power product with the same base), verify the correct application of the corresponding property based on the explicit result.	Remembering the properties of the powers. Working with expressions that contain powers.	Analysing expressions with properties of powers, identifying: the use or not of the properties of the powers; the correct application of the property.
1	Indicate any correctness, by clicking on one of the two alternative options presented; write the operation that must be carried out among the exponents; identify and write the base and the exponent of the produced power; write the resulting power.	Knowing the concept of power at full exponent, identifying the base and the exponent; remembering the property of the product of similar powers; calculating the sum between positive integers.	Knowing and knowing how to apply theorem about the product of similar powers.
2	Given an equality of whether it is correct or not, enunciate a property of the powers, recognize the base and the exponent of a power, and apply a property of the powers.	Knowing the properties of the powers, knowing the meaning of the symbol =.	Recognizing the correctness or of the application of a property of the powers.
3	In my opinion the O-P-R descriptors are well explained, I would just have used the verb 'select' instead of 'verify' in the performance.		

5.4 Final Questionnaire

Instructors appreciated several aspects of the activity: 'Clarity of explanations', 'Adequacy of the topics covered', 'Completeness of the topics covered', 'Course progress modality' were rated more than 4.5 out of 5 on average (with standard deviation of less than 0.75). They showed high appreciation of the use of P-R-O triad in the activity: similar rates were obtained for the following aspects:
- Designing learning units through the specification of the P-R-O descriptor triad;
- Having the three-part P-R-O descriptors associated with learning units of other teachers available;
- Reasoning together with peers about learning units on the basis of the P-R-O modeling principles.

Overall, instructors believe that this model will be useful. The use of the P-R-O descriptors may be useful to reflect systematically on teaching, to elaborate material for students, to activate discussions within the Community, to share material in the Community, and to search for material. Those aspects were rated 4.16 or more on average (with standard deviation less than 0.93).

The descriptors were used by instructors during the phases of planning, realization and use of a question, as highlighted by the high average rates of Table 2.

Table 2. Average rates of descriptors appreciation considering three phases

	Design	Realization	Use
The use of the descriptor Performance helps reflect on how the question allows to measure the achievement of the intended goals	4.30 (0.80)	4.22 (0.91)	4.17 (0.77)
The use of the descriptor Requirements helps reflect on the requirements necessary to answer the question	4.63 (0.58)	4.33 (0.67)	3.70 (1.09)
The use of the descriptor Objectives helps reflect on the learning objectives	4.44 (0.75)	4.31 (0.74)	3.85 (0.86)

Open answers confirm both the appreciation of the overall activity and the expectation that the systematic adoption of the P-R-O model would be helpful to the Community. A selection of answers to the question "Indicate how you plan to use what was proposed during experimentation in your teaching" follow.

- **Helpfulness for the new instructors joining the community**: "Having just entered the PPS community, I have not yet mastered the use of the Automatic Assessment System, so I am not yet able to prepare ex novo questions; I would like to start using the questions in the archive and therefore I think it would be very useful if all the questions were accompanied by the proposed descriptors".
- **Alignment to everyday teaching practice**: "During the experimentation I realized that before formulating a test, I always reflected on objectives, performance and requirements, but I never explicitly wrote them. Now I will begin to do it, because I think it is essential to draw a learning path. I also think it is important to communicate it to the students so that they too have clear that each trial has its own purpose and is part of a larger project".
- **Positive effects of cooperation between peers**: "The questions proposed by our colleagues have been a useful tool, because they allow us to understand not only what are the objectives that guide us to create a test, but also how important it is to be clear and precise in the delivery we propose".

All the participants would like to participate to a second experimentation to further discuss and learn about the proposed model. Among the suggestions left through open answers, they indicated useful hints for future experimentations:

- having a list of counterexamples, that are incorrect associations of descriptors to materials;
- repeating, strengthening and supporting the activity with a supervising tutor;
- systematizing descriptors cross-review, to make them as objective and recognizable as possible;
- using a standardized vocabulary, so that terms are not subject to individual interpretations.

6. CONCLUSION

Associating descriptors to materials should increase the educational validity of the whole instructors' work, from the design phase to the sharing and finding of different resources. It could especially favour the selection of materials by instructors who join Communities for the first time, as they might not be familiar with the tools. Furthermore, clarifying learning intentions is crucial for interpreting students' results.

The main outcome of the experimentation is the creation of the first collection of shared resources associated with descriptors validated by experts: the elaborated materials and the guidelines to their preparation are made available to all the instructors in order to be used in their own didactics. Furthermore, workshop submissions will be reviewed in order to create a gold standard of resources described and counter examples with comments for the teachers who want to continue using this model. Reviewing instructors' work is also strategic for a careful development of the ontologies and the planning of the system's implementation. A web-based software will be integrated into Learning Management Systems hosting targeted Communities for the experimentation of the proposed model, so that experimentation involving students will enable to further investigate material authoring and automatic dispatching.

ACKNOWLEDGEMENT

We sincerely thank the teachers involved in the experimentation and the Scientific Committee of PP&S Project.

REFERENCES

Barana, A. et al, 2015. Automated Assessment in Mathematics. *Proceedings of 2015 IEEE 39th Annual International Computers, Software & Applications Conference*. Taichung, Taiwan, pp. 670-671.

Ravotto, P., 2011. Competence-based learning in Europe & the sloop2desc MODEL. *In Preparing the teachers for a competence-based education system*, pp.13-16.

Trinchero, R., 2017. *Costruire e certificare competenze nel secondo ciclo*. Rizzoli Education, Milano, Italy.

Vincent, T. et al, 2014. A new framework of concept clustering and learning path optimization to develop the next-generation e-learning systems. *In Journal of Computers in Education*, pp. 335-352.

Huang, C. et al, 2015. Constructing Learning Maps for Lecture Videos by Exploring Wikipedia Knowledge, *Proceedings (Part I) of 16th Pacific-Rim Conference on Multimedia*, Gwangju, South Korea, pp. 559-569.

Winterton, J. et al, 2006. *Typology of knowledge, skills and competences : clarification of the concept and prototype*. Office for Official Publications of the European Communities, Luxembourg.

Black, P. and Wiliam, D., 2009. Developing theory of formative assessment. *In Educational Assessment, Evaluation and Accountability*, Vol. 21, No. 1, pp. 5-31.

Wiliam, D. and Thompson, M., 2007. Integrating assessment with instruction: What will it take to make it work?. *In The future of assessment: Shaping teaching and learning*, pp. 53-82.

Di Caro, L. et al, 2018a. Alignment of content, prerequisites and educational objectives: towards automated mapping of digital learning resources. *Proceedings of The 14th International Scientific Conference eLearning and Software for Education*. Bucharest, Romania, pp. 335-342.

Barana, A. et al, 2018a. Developing Competence Assessment Systems in e-Learning Communities. *Proceedings of EDEN 2018 ANNUAL Conference*. Genova, Italy, pp. 879-888.

Di Caro L. et al, 2018b. Towards adaptive systems for automatic formative assessment in virtual learning communities. *Proceedings of 2018 42nd IEEE International Conference on Computer Software & Applications*. Tokyo, Japan, pp. 1000-1005.

Barana, A. and Marchisio, M., 2016. Ten good reasons to adopt an automated formative assessment model for learning and teaching Mathematics and scientific disciplines. *In PROCEDIA: SOCIAL & BEHAVIORAL SCIENCES*, Vol. 228, pp. 608-613.

Barana, A. et al, 2018b. Sharing system of learning resources for adaptive strategies of scholastic remedial intervention. *Proceedings of 4rd International Conference on Higher Education Advances (HEAd'18)*. Valencia, Spain, pp. 1465-1473.

Ellerani, F. and Parricchi, M., 2010. *Ambienti per lo sviluppo professionale degli insegnanti. Web 2.0, gruppo, comunità di apprendimento*. Francoangeli, Milano, Italy.

Demartini, C. et al, 2013. PP&S100 - una comunità di comunità di collaborative learning attraverso le nuove tecnologie. *Proceedings of DIDAMATICA 2013, Tecnologie e Metodi per la Didattica del Futuro*. Pisa, Italy, pp. 989-998.

Brancaccio, A. et al, 2014. Il progetto PP&S informatica a scuola. *In MONDO* DIGITALE, Vol. 13, No. 51, pp. 565-574.

Brancaccio, A. et al, 2015. Problem Posing and Solving: Strategic Italian Key Action to Enhance Teaching and Learning of Mathematics and Informatics in High School. *Proceedings of 2015 IEEE 39th Annual International Computers, Software & Applications Conference*. Taichung, Taiwan, pp. 73-88.

Brancaccio, A. et al, 2016. L'efficacia dell'apprendimento in rete degli immigrati digitali. L'esperienza SMART per le discipline scientifiche. *In MONDO DIGITALE*, Vol. 15, No. 64, pp. 803-821.

Trinchero, R., 2014. *Valutare l'apprendimento nell'e-learning*. Erickson, Trento, Italy.

Anderson, L.W. et al, 2001. *A taxonomy for learning, teaching, and assessing. A revision of Bloom's taxonomy of educational objectives*. Addison Wesley Longman, New York.

Nevzorova, O. et al, 2014. OntoMathPRO ontology: A linked data hub for mathematics. *Proceedings of KESW: International Conference on Knowledge Engineering and the Semantic Web*. Kazan, Russia, pp. 105-119.