

## Automatic Evaluation for E-Learning Using Latent Semantic Analysis: A Use Case

Mireia Farrús et Marta R. Costa-jussà

Volume 14, numéro 1, mars 2013

URI : <https://id.erudit.org/iderudit/1066981ar>  
DOI : <https://doi.org/10.19173/irrodl.v14i1.1389>

[Aller au sommaire du numéro](#)

Éditeur(s)

Athabasca University Press (AU Press)

ISSN

1492-3831 (numérique)

[Découvrir la revue](#)

Citer cet article

Farrús, M. & R. Costa-jussà, M. (2013). Automatic Evaluation for E-Learning Using Latent Semantic Analysis: A Use Case. *International Review of Research in Open and Distributed Learning*, 14(1), 239–254.  
<https://doi.org/10.19173/irrodl.v14i1.1389>

Résumé de l'article

Assessment in education allows for obtaining, organizing and presenting information about how much and how well the student is learning. The current paper aims at analysing and discussing some of the most state-of-the-art assessment systems in education. Later, this work presents a specific use case developed for the Universitat Oberta de Catalunya, which is an on-line University. An automatic evaluation tool is proposed so that it allows the student to evaluate himself anytime and receive instant feedback. This tool is a web-based platform and it has been designed for engineering subjects (i.e. with math symbols and formulas) in Catalan and Spanish. Particularly, the technique used for automatic assessment is Latent Semantic Analysis. Although the experimental framework from the use case is quite challenging, results are promising.

Copyright (c) Mireia Farrús, Marta R. Costa-jussà, 2013



Ce document est protégé par la loi sur le droit d'auteur. L'utilisation des services d'Érudit (y compris la reproduction) est assujettie à sa politique d'utilisation que vous pouvez consulter en ligne.

<https://apropos.erudit.org/fr/usagers/politique-dutilisation/>

**é**rudit

Cet article est diffusé et préservé par Érudit.

Érudit est un consortium interuniversitaire sans but lucratif composé de l'Université de Montréal, l'Université Laval et l'Université du Québec à Montréal. Il a pour mission la promotion et la valorisation de la recherche.

<https://www.erudit.org/fr/>

## Automatic Evaluation for E-Learning Using Latent Semantic Analysis : A Use Case



Mireia Farrús  
Universitat Pompeu Fabra, Spain

Marta R. Costa-jussà  
Institute for Infocomm Research, Singapore

### Abstract

Assessment in education allows for obtaining, organizing, and presenting information about how much and how well the student is learning. The current paper aims at analysing and discussing some of the most state-of-the-art assessment systems in education. Later, this work presents a specific use case developed for the Universitat Oberta de Catalunya, which is an online university. An automatic evaluation tool is proposed that allows the student to evaluate himself anytime and receive instant feedback. This tool is a web-based platform, and it has been designed for engineering subjects (i.e., with math symbols and formulas) in Catalan and Spanish. Particularly, the technique used for automatic assessment is latent semantic analysis. Although the experimental framework from the use case is quite challenging, results are promising.

**Keywords:** E-learning; automatic test assessment; web platform; latent semantic analysis

## Introduction

Assessment in education is the process of obtaining, organizing, and presenting information about what and how the student is learning. Assessment uses several techniques during the teaching-learning process, and it is especially useful when evaluating open-answer questions since they allow teachers to better understand the assimilation of the student in the subject. In some cases, for instance, students with high punctuation in closed-answer tests report subjacent conceptual errors when being interviewed by a teacher (Tyner, 1999).

During the last years, the use of a computer for assessment purposes has substantially increased. The aims of using computer assessment include achieving and consolidating the advantages of a system with the following characteristics (Brown et al., 1999): first, to reduce the professors' workload by automating part of the student evaluation task; second, to provide the students with detailed information on their learning period in a more efficient way than traditional evaluation; and, finally, to integrate the assessment culture into the students' daily work in an e-learning environment. In fact, nowadays one of the most crucial things in assessment is feedback, so assessment of learning is generally intended to measure learning outcomes and report those outcomes to students (and not only to the system or teacher).

The current paper aims at analysing some of the most state-of-the-art assessment systems in education and presents a specific use case developed for the Universitat Oberta de Catalunya. Some examples of existing e-learning platforms are given. Next the use of latent semantic analysis as a semantic analyser algorithm of related documents is briefly described and explained in the context of assessment tasks. Then the authors present the above-mentioned use case, which takes advantage of latent semantic analysis in order to obtain the evaluation results. Finally, conclusions are shown.

## E-learning Assessment Platforms

Some papers in the literature are oriented to automated essay-scoring research. The most relevant ones can be found in Miller (2003), Shermis and Burstein (2003), Hidekatsu et al. (2007), and Hussein (2008). However, studies covering automatic essay scoring in engineering subjects are limited (to the best of our knowledge), though not inexistent. In Quah et al. (2009), for instance, the authors use a Support Vector Machine to build a prototype system, which is able to evaluate equations and short answers. The system extracts textual and mathematical data from input files in the form of distinct words for text and for mathematical equations using equation trees based on MathTree format. Then the system learns how to evaluate them, based on grades given at the beginning, learning the evaluation scheme and evaluating the subsequent scripts automatically.

Many portals can be currently found online. To overview some examples, for instance, the Online Learning and Collaboration Services (OLCS, <http://www.olcs.lt.vt.edu>) from VirginiaTech provides system administration, support, and training for scholars, online course evaluations, and other instructional software. The ViLLE Collaborative Educational Tool (<http://ville.cs.utu.fi/>) is a full environment capable of doing many kinds of assessment, where people can benefit of developing their own material instead of developing a new Web site. In addition, it becomes easier to get feedback on the material if done in collaboration with other teachers.

Another example of a learning platform is the Khan Academy (<http://www.khanacademy.org>), which has created a generic framework for building exercises. This framework, together with the exercises themselves, can be used completely independently of the Khan Academy application. The framework exists in two components: an HTML markup for specifying exercises and a plug-in for generating a usable and interactive exercise from the HTML markup.

Furthermore, some systems can be found specifically for math exercises. STACK (<http://www.stack.bham.ac.uk>), for instance, is an open-source system for computer-aided assessment in mathematics and related disciplines, with emphasis on formative assessment. And some systems such as restructured text (<http://docutils.sourceforge.net/rst.html>) provide techniques that can be used to develop new materials.

## Latent Semantic Analysis in E-Learning

The task of evaluating a document in our education context implies judging the semantic content of such a document. To this end, latent semantic analysis (LSA), also known as latent semantic indexing, a technique that analyses a semantic relationship between a set of documents and the terms they contain (Hofmann, 1999), has been successfully applied in multiple natural language processing areas such as cross-language information retrieval (Dumais et al. 1996), cross-language sentence matching (Banchs & Costa-jussà, 2010), and statistical machine translation (Banchs & Costa-jussà, 2011).

The aim of LSA is to analyse documents in order to find their underlying meaning or concepts. The technique arises from the problem of how to compare *words* to find relevant documents since what we actually want to do is compare concepts and meanings that are behind the words, instead of the words themselves. In LSA, both words and documents are mapped into a *concept space*. It is in this space where the comparison is performed. This space is created by means of the well-known singular value decomposition (SVD) technique, which is a factorization of a real or a complex matrix (Greenacre, 2011).

In the specific area of essay assessment, LSA has shown promising results in content analysis of essays (Landauer et al., 1997), where LSA-based measures were closely related to human judgments in predicting how much the student will learn from the text (Wolfe et al., 2000; Rehder, et al., 2000) and in grading essay answers (Kakkakonen et al., 2005). Other educational applications are intelligent tutoring systems which provide help for students (Wiemer- Hastings et al., 1999, Foltz et al., 1999b) and assessment of summaries (Steinhart, 2000). In this context, LSA has been applied to a variety of languages such as essays written in English (Wiemer-Hastings & Graesser, 2000), in French (Lemaire and Dessus, 2001), and in Finnish (Kakkakonen et al, 2005) since LSA is language independent. All these studies show that, although it does not take into account word ordering, LSA is capable of capturing significant portions of the meaning not only of individual words but also of whole passages such as sentences, paragraphs, and short essays. That is why we have chosen LSA in order to compare the semantic similarity of documents in the concept space (Pérez et al., 2006).

Particularly, in this work and differently from the previous literature, we investigate if LSA can be applied for e-assessment of mathematical essays. Additionally, experiments are performed both in Catalan and Spanish. LSA is integrated as follows. The documents containing the responses of the students are compared with one or more reference documents containing the correct answers created by the teachers. Then such semantic comparison of the students' and reference documents will allow teachers to generate an approximate evaluation of the students. For the document comparison and/or document retrieval, documents are typically transformed into a suitable representation, usually a vector-space model (Salton, 1989). A document is represented as a vector, in which each dimension corresponds to a separate term. If a term occurs in the document, its value in the vector is non-zero. Several ways of computing these values, also known as (term) weights, have been developed. One of the best known schemes is tf-idf (term frequency inverse document frequency) weighting. The tf-idf weight defines statistically how important a word is to a document in a collection. Such a representation is known to be noisy and sparse. That is why in order to obtain more efficient vector-space representations, space reduction techniques are applied (Deerwester et al., 1990; Hofmann, 1999; Sebastiani, 2002), so that the new reduced space is supposed to capture semantic relations among the documents in the collection. Figure 1 shows a schematic representation of the use of latent semantic analysis for automatic essay scoring.

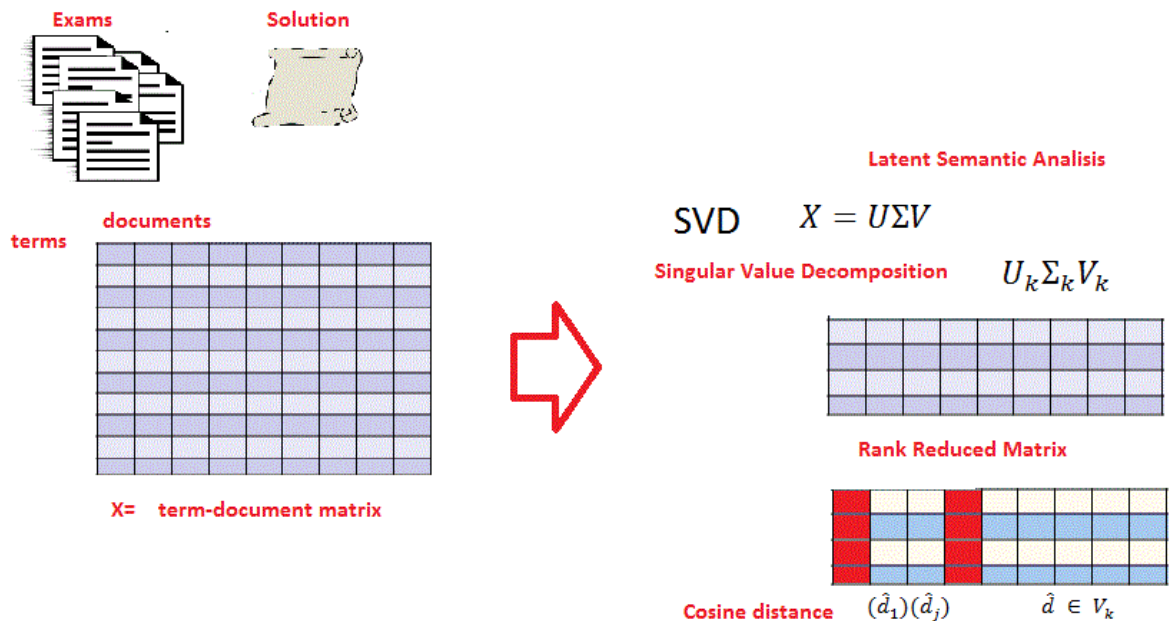


Figure 1. Schematic representation of the use of latent semantic analysis for automatic essay scoring is the term-document matrix and is the singular value decomposition of the matrix, which allows computing a rank reduction matrix over which the cosine distance among documents is computed.

As a final step, a cosine distance similarity measure among each exam and its solution in the reduced space is calculated, obtaining a score that shows how a particular set of exams is similar in semantics with their corresponding solution.

## The UOC's Use Case

This section addresses the creation of a free-text assessment tool through the Internet, allowing the automatic student assessment of the Universitat Oberta de Catalunya (Open University of Catalonia, UOC). The main characteristics of the university assessment system and the developed tool are described in the following subsections.

### The Universitat Oberta de Catalunya

The UOC is an online university based in Barcelona with more than 54,000 students. Over 2,000 tutors and faculty work together, and administrative staff of around 500 provide services to all these students. The students follow a continuous assessment system, which is carried out online throughout the semester. Although this system is successfully used to complete their studies, one of the main problems is the growing number of students each year, which makes the task of marking their continuous assessment tedious and time-consuming. Likewise, more external tutors are needed to carry out this task, which makes it difficult to come to agreement on criteria.

## The Assessment Tool

The tool developed at the UOC aims to provide an automatic assessment of assignments in the engineering subjects by using the latent semantic analysis technique, following the work carried out by Miller (2003), where the application of LSA to automated essay scoring is examined and compared to earlier statistical methods for assessing essay quality. The implementation of LSA is done using JAVA.

The web-based free-text assessment tool allows the professors to design as many evaluation tests as they want, with as many questions as they consider necessary for the evaluation. On the one hand, for each question, the professor associates several correct-answer models in order to generate enough reference answers to guarantee that the automatic evaluation system works correctly. On the other hand, the web-based platform allows students to realise as many evaluation tests as they want, generating, after each test realization, a report including the evaluation results of every individual question as well as the overall results. Moreover, the tool provides the students with the possibility of comparing the reference answers generated by the professor with their own answers in order to give detailed feedback and improve their learning process. The platform also includes a text editor that allows inserting formulas both in the statements and in the answers with the JavaScript plug-in MathML (Su et al., 2006).

## Evaluation Experiments

In this section we describe the experimental framework in our case study. We include subsections that particularly describe the working framework, the web interface, the assessment experiments, and the results obtained.

### **Working framework.**

The main objective of the tool is to help teachers in their evaluation tasks on a large number of students. These first experiments involve a controlled and relatively small number of students in order to establish the groundwork for further and more extensive experiments. The application framework covers the students in two consecutive semesters (with 54 and 70 registered students, respectively) of a single UOC's subject called *Circuit Theory*, a core subject belonging to the first year of UOC's Telecommunications Engineering Grade.

Apart from the single final evaluation that takes place at the end of the semester, the subject's assessment model contains four different single continuous assessment assignments (CAAs) distributed over the course of the semester and a single practical work that includes computer simulation exercises, structured as follows. The first three CAAs are made up of two different sections: a short question section and an exercises section. The fourth and last CAA contains only an exercises section. More specifically, the short question sections consist of a set of 5-6 questions about very concrete issues. Each of these questions is provided with four possible answers, where only one of them



is correct, in such a way that the students have to specify the correct answer and give reasons for their choices. Due to the technical nature of the subject matter, mathematical equations usually appear in the wording of both questions and answers as well as in the students' corresponding justifications.

Within this context, the short questions section of the first three CAAs have been chosen as a specific application framework to perform the automatic evaluation experiments, due to the suitability of the structure and length of both the question and answers as well as to the nature (short text plus a few mathematical equations) of the justifications the students have to provide.

### Web interface.

The automatic test assessment system is presented as a web platform, where access can be realized from two different profiles: the teacher and the student. The main task of the teacher is to provide questions and correct reference answers. Thus, a teacher can realise two different actions for each subject: to create a new test and to modify an existing one. In order to create a new test, the teacher must first define the following attributes: the name of the test, the subject in which it belongs, the position within the test set of the subject, and a brief description (Figure 2a). Once these attributes have been inserted, the teacher can register the empty test in the database. Then teachers can insert as many questions as they wish in the test. For each new question, the following attributes need to be completed: (a) statement, (b) maximum possible mark (c) minimum mark to pass the question, (d) question difficulty, and (e) language of the statement (Figure 2b). Moreover, a set of reference answers is associated with each question. Additionally, the teacher can consult the obtained results as well as the answers given by the students.

Figure 2 consists of two screenshots of a web interface. Screenshot (a) is titled 'Introducir Nuevo Test' and shows a form with the following fields: 'Asignatura' (Procesamiento del Lenguaje Natural I), 'Nombre del Test' (Lenguajes y Gramáticas), 'Numero de Test' (1), and 'Descripcion (255 caracteres)' (Un pequeño repaso a los conomientos de los lenguajes y gramáticas visto en clase). A 'registrar' button is at the bottom. Screenshot (b) is titled 'Introducir Nueva Pregunta' and shows a form with the following fields: 'Asignatura' (Procesamiento del Lenguaje Natural I), 'Test' (Lenguajes y Gramáticas), 'Enunciado de la Pregunta (255 caracteres)' (a large text area with a rich text editor toolbar), 'Nota Maxima [0-10]' (10), 'Nota de Corte [0-10]' (5.35), 'Dificultad' (facil), and 'Idioma' (castellano). A 'registrar' button is at the bottom.

Figure 2. Creation page of a new test (a) and creation form of a new question (b).



Once authenticated, the students can perform the following actions: (1) evaluating themselves by realising a test, (2) checking the history of the realised tests, and (3) consulting the obtained marks as well as the maximum and minimum marks defined by the teacher.

In order to evaluate themselves, students are shown a list of alphabetically ordered subjects in which they can realise the evaluation by choosing a subject and selecting the test they wish to start with and the difficulty level. The statement of each question is presented to the students together with their corresponding mark. The students must answer within a text editor, where they can insert formulas thanks to a JavaScript plug-in called MathEdit (Su et al., 2006), as seen in Figure 3a. Once the answer has been written and the test is finished, the system provides a score to the student together with the obtained marks in each of the questions (see Figure 3b). Likewise, the students can check, for each question, the answers they wrote as well as the reference questions written by the teacher.



Figure 3. Question and text editor with MathEdit (a) and mark of the test once it is finished (b).

Apart from the realisation of the tests, the students have the possibility of logging into the platform in order to evaluate their progress. Thus, every student has access to a history in which they can see a list of completed tests. Once a completed test is chosen, the questions can be seen in detail, including the answer given by the student, the obtained mark, the maximum and minimum marks defined by the teacher, and the reference answers used by the automatic evaluation system in order to make the corrections.

### Assessment experiments.

This section describes the automatic evaluation performed over the continuous assessment assignments of the students. The experiments carried out used the CAAs from two consecutive semesters, S1 and S2, in which 54 and 70 students were

registered, respectively. Each semester included a set of three different CAAs (CAA1, CAA2, and CAA3). The data were tokenized, lowercased. The 20 most frequent words were discarded. As follows, we describe the procedure for treating the set of solutions with LSA:

1. Compute N solutions in terms of tf-idf:
  - a) Extract vocabulary
  - b) Each solution is a vector of M dimensions
2. Matrix solution  $N \times M$
3. Compute SVD
4. Select L singular values

Then, for each student answer the procedure is as follows.

1. Vectorise the answer in terms of tf-idf, use the vocabulary of the set of solutions. We've got a vector of dimension M.
2. Project the vector into the reduced space.
3. Compute the similarity of this reduced space vector with each solution. We keep the maximum distance.

The material used in the analysis presented three main problems.

1. Format files. The students' CAAs are delivered in many different formats, although they are mainly in PDF, Word, and Open Office Writer. Some of them are even scanned documents pasted as image files in Word or Writer documents. Therefore, not all the CAAs can be easily transformed into TXT format to be treated properly. Consequently PDF documents and all those documents containing image files were removed from the original set of files. Table 1 shows, for each semester, the number of registered students, the number of original documents, and the number of used documents after removing PDF documents and documents with pasted images. The table also shows the vocabulary for each CAA. As can be seen, the vocabulary size is not correlated with the number of CAAs, so the vocabulary content of the CAAs varies largely among each set.
2. Mathematical formulation. Given that we are using a *bag-of-words* approach, the formulation extracted from Open Office documents was coded in MathML (Mathematical Markup Language), while the formulation extracted from Word documents was not, which made a big difference between CAAs regarding the final vocabulary.

3. **Language.** The students submitted the CAAs in both the Catalan and Spanish languages. In this case, we assumed that the method presented in the current paper is able to take advantage of the vocabulary that is language independent, such as the mathematical variables.

Table 1

*Registered Students, Number of Original CAAs (#orig.), Number of Used CAAs (#used), and Vocabulary Size Used (vocab.) for each Semester*

| Semester | Students | CAA1   |       |        | CAA2   |       |        | CAA3   |       |        |
|----------|----------|--------|-------|--------|--------|-------|--------|--------|-------|--------|
|          |          | #orig. | #used | vocab. | #orig. | #used | vocab. | #orig. | #used | vocab. |
| S1       | 54       | 20     | 14    | 857    | 19     | 13    | 730    | 15     | 10    | 712    |
| S2       | 70       | 28     | 20    | 1027   | 25     | 9     | 699    | 20     | 16    | 1291   |

### Results.

In order to carry out the preliminary assessment experiments, CAA1 and CAA2 from semester S1 were used as development material, which allowed concluding that the best rank reduction in latent semantic analysis was five.

The results are shown in terms of the correlation obtained between automatic and human evaluations. We define *human evaluation* as the assessment made by the teacher in a traditional way, while *automatic evaluation* is defined as a computer-based assessment given by the methodology proposed in the current work (i.e., the quantifications obtained automatically using latent semantic analysis and the cosine distance).

Thus, by using the latent semantic analysis, automatic evaluations were obtained for each student, CAA, and semester. Then the correlations between automatic and human evaluations were computed for each semester and CAA collection. The correlation results obtained are reported in Table 2 (*correlation* column), together with the statistical significance of the correlation results (*p* column).

As can be seen from the table, in statistically significant results (i.e., where  $p < 0.05$ ), the correlation varies from 52% to 69% (see CAA1 and CAA2 from semester S2). Although these results are lower than those presented in Miller (2003), they are promising given that we are dealing with a complete textual subject, but with a subject

containing a considerable number of mathematical formulas. The rest of the results (S1 and CAA3 from S2) are not statistically significant.

On the one hand, we must take into account that the reference answers were written in Catalan by the teachers, while the students could choose whether to answer the tests in Catalan or Spanish, so the language of the tests was not the same in all the students' CAAs. On the other hand, unlike the students' CAAs, all the reference solutions were available in Writer format. Since only the mathematical formulas of the Writer documents were transformed into MathML, there was also disparity in the formulas in each CAA collection.

In order to see how these disparities could have affected the results, we computed the percentage of CAAs in each set that satisfied the following two requirements at the same time (i.e., the same two characteristics satisfied by the reference solutions).

1. The formulas were coded in MathML.
2. The students answered in the Catalan language.

The percentage of CAAs satisfying both characteristics are shown in Table 2 in the third column of every CAA result. It can be seen that the two statistically significant results with a correlation over 50% (i.e., CAA1 and CAA2 from semester S2) correspond to those results in which the codification and the language used is the same as the reference solutions in more than 25% of the cases. Therefore, it could be stated from the results that the correlation between human and automatic evaluations depends on the coherence of both the mathematical codification and the language used in the tests.

Table 2

*Correlation Results (corr.) and Statistical Significance (p) between Automatic and Human Evaluation, and Percentage of CAAs Satisfying the Same Characteristics as the Reference Solutions (same charact.)*

| Semester | CAA1  |          |               | CAA2  |          |               | CAA3  |          |               |
|----------|-------|----------|---------------|-------|----------|---------------|-------|----------|---------------|
|          | corr. | <i>p</i> | same charact. | corr. | <i>p</i> | same charact. | corr. | <i>p</i> | same charact. |
| S1       | 16%   | 0.60     | 14%           | 12%   | 0.68     | 15%           | 15%   | 0.68     | 10%           |
| S2       | 52%   | 0.04     | 30%           | 69%   | 0.04     | 28%           | 29%   | 0.27     | 25%           |

For example, from CAA1 and S1, one answer to a short question to be evaluated was, “Si introduïm un senyal sinusoidal en un circuit, la resposta forçada serà una sinusoide que l’entrada amplificada per  $H(s)$ ” (in English, “If we introduce a sinusoidal signal in a circuit, the forced response is a sinusoid amplified by the input  $H(s)$ ”). The answer was, “La resposta del sistema és una senyal sinusoidal de la mateixa freqüència amplificada per  $H(s)$ ” (in English, “The system response is a sinusoidal signal of the same frequency amplified by  $H(s)$ ”). There is only a detail de la mateixa freqüència (in English, the same frequency) which is not present in the student answer. This answer is ranked by the teacher as an 8 and by the system as a 9.

To conclude the presented results, it may be interesting to discuss briefly the role played by MathML, as opposed to the words in the written reports. At the time of realising the current experiments, mathematical formulas were merely treated as words. In fact, one of the drawbacks of the current study is that we are dealing with the bag of words method; therefore, the word order, which is definitely important in the meaning of mathematical formulas, is not taken into account. For instance, the method does not distinguish between  $I=V/R$  and  $I=R/V$ . However, since the former is totally correct, the latter is completely wrong. This is one of the challenges to be solved in future research.

## Conclusions

This paper has presented an analysis and a discussion of state-of-the-art assessment systems in education. Additionally, this work shows a detailed case study of an automatic correction tool embedded as part of virtual classrooms in UOC’s web-based teaching-learning environment in order to help students’ self-assessment by providing them with instant feedback. Thereby, adult e-learners, who usually have a lack of time, do not have to await teachers’ assessments to be graded. This tool, based on a web interface is designed to be used in an online environment, both by the teacher (the correct design and assessment tests) and student (the self-assessed). The automatic evaluation process is based on testing techniques using natural language processing and latent semantic processing.

The case study carried out in this paper has had to overcome some problems regarding the available material, first of which is the existence of a lot of mathematical formulas in the engineering subjects treated. Although many research works have dealt with automated essay scoring, as far as we are concerned, they have not dealt with mathematical language. Moreover, the students’ tests are available in different languages and file formats, which makes it even more difficult to treat the mathematical formulas by converting them into a homogeneous code.

In order to be able to treat the available material, PDF documents and those Word or Writer documents containing pasted images as responses were removed at the

beginning. However, we are aware that this is not the best method to collect the data, and both of them (PDF and image files) will be dealt with in future research.

Nevertheless, despite the difficulties in the material used, the preliminary experiments have shown some interesting results. After computing the correlation between the automatic and the human assessment tests it was shown that only two from the six evaluation tests provided correlation greater than 50% with statistically significant results. These two sets correspond to those set of PACs that have more similarity with the reference solution PACs: The mathematical formulas are coded in MathML, and the students answers were mostly written in the same language.

In automatic essay assessment we would expect a higher correlation. However, we are dealing with a challenging issue since it does include mathematical symbols and formulas, which makes the current analysis more difficult. Therefore, although for the time being the correlation results are not satisfactory, they have set a starting point that allows us to work with this kind of material in engineering subjects. Thus, future work will focus on improving the format of the materials to give coherence to them (i.e., by using the same formulation and dealing with the language issue). Additionally, we plan to experiment with non-linear space reduction such as multidimensional scalability in order to find further semantic similarities.

## Acknowledgements

The authors would like to thank the Universitat Oberta de Catalunya for providing us the materials and the context needed to develop the current research, and for partially funding this work under the Teaching Innovation Project number IN-PID 1043. We would like to specially thank Germán Cobo, David García, Jordi Duran, Francisco Cortés, Lluís Villarejo, and Rafael E. Banchs for their support to this work. This work has also been partially funded by the Seventh Framework Program of the European Commission through the International Outgoing Fellowship Marie Curie Action (IMTraP-2011-29951).

## References

- Banchs, R. E., & Costa-jussà, M. R. (2010). A non-linear semantic mapping technique for cross-language sentence matching. *Lecture Notes in Computer Science, Springer, 7th International Conference on Natural Language Processing (IceTAL)* (pp. 57-66). Iceland.
- Banchs, R. E., & Costa-jussà, M. R. (2011, June). *A semantic feature for statistical machine translation*. ACL HLT: 5th Workshop on Syntax, Semantics and Structure in Statistical Translation (SSST-5), Portland.
- Brown, S., Race, R., & Bull, J. (1999). *Computer-assisted assessment in higher education*. Kogan Page.
- Deerwester, S., Dumanis, S., Furnas, G.W., Landauer, T.K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science, 41*(6), 391-407.
- Dumais, S., Landauer, T., & Littman, M. (1996). Automatic cross-linguistic information retrieval using latent semantic indexing. In *SIGIR 1996 Workshop on Cross-Lingual Information Retrieval*.
- Foltz, P. W., Laham, D., & Landauer, T. K. (1999b). The intelligent essay assessor: Applications to educational technology. *Interactive Multimedia Electronic J. of Computer-Enhanced Learning*. Retrieved from <http://imej.wfu.edu/articles/1999/2/04/index.asp>
- Greenacre, M. (2011). *It had to be U – The SVD song*. Available from [http://www.youtube.com/watch?v=JEYLFIVvR9I&feature=player\\_detailpage](http://www.youtube.com/watch?v=JEYLFIVvR9I&feature=player_detailpage).
- Hidekatsu, K., Kiyoshi, A., Chiharu, I., Nagatomo, N., & Shinya, W. (2007). Toward a software development model for automatic marking software. *Proceedings of the 35<sup>th</sup> Annual ACM SIGUCCS Conference on User Services*. Orlando, Florida, USA.
- Hofmann (1999). Probabilistic latent semantic analysis. *Proceedings of Uncertainty in Artificial Intelligence, UAI99I* (pp. 289-296).
- Hussein, S. (2008). Automatic marking with Sakai. *Proceedings of the 2008 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists on IT Research in Developing Countries: Riding the Wave of Technology*. Wilderness, South Africa.
- Landauer, T. K., Laham, D., Rehder, B., & Schreiner, M. E. (1997). How well can passage meaning be derived without using word order? A comparison of latent semantic



- analysis and humans. In *Proceedings of the 19th Annual Meeting of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.
- Lemaire, B., & Dessus, P. (2001). A system to assess the semantic content of student essays. *Journal of Educational Computing Research*, 24, 305–320.
- Miller, T. (2003). Essay assessment with latent semantic analysis. *Journal of Educational Computing Research*, 29(4), 495-512.
- Pérez, D., Alfonseca, E., Rodríguez, P., & Pascual, I. (2006). Willow: Automatic and adaptive assessment of students free-text answers. In *Proceedings of the 22nd International Conference of the Spanish Society for the Natural Language Processing (SEPLN)*.
- Quah, J.T-S., Lim, L-R., Budi, H., & Lua, K-T. (2009). Towards automated assessment of engineering assignments. *Proceedings of the 2009 International Joint Conference on Neural Networks* (pp. 1411-1418). Atlanta, Georgia, USA.
- Rehder, B., Schreiner, M. E., Wolfe, M. B., Laham, D. Landauer, T. K., & Kintsch, W. (1998). Using latent semantic analysis to assess knowledge: Some technical considerations. *Discourse Processes*, 25, 337-354.
- Salton, G. (1989). *Automatic text processing: The transformation, analysis and retrieval of information by computer*. Addison-Wesley.
- Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Computing Surveys*, 34(1), 1–47.
- Shermis, M. D., & Burstein, J. C. (2003). *Automated essay scoring: A cross-disciplinary perspective*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Steinhart, D. (2000). *Summary street: An LSA based intelligent tutoring system for writing and revising summaries* (Ph.D. thesis). University of Colorado, Boulder, Colorado.
- Su, W., Wang, P., Li, L., Li, G., Zhao, Y. (2006). MathEdit, a browser-based visual mathematics expression editor. *Proceedings of The 11th Asian Technology Conference in Mathematics* (pp. 271-279.). Hong Kong.
- Tyner, K. (1999). *Development of mental representation: Theories and applications*. Lawrence Erlbaum Associates.
- Wiemer-Hastings, P., Wiemer-Hastings, K., & Graesser, A. (1999). Approximate natural language understanding for an intelligent tutor. In *Proc. of the 12th Int'l Artificial Intelligence Research Symposium* (pp. 172–176). Menlo Park, CA, USA.

- Wiemer-Hastings, P., & Graesser, A. (2000). Select-a-kibitzer: A computer tool that gives meaningful feed-back on student compositions. *Interactive Learning Environments*, 8, 49–169.
- Wolfe, M. B., Schreiner, M. E., Rehder, B., Laham, D., Foltz, P. W., Kintsch, W., & Landauer, T. K. (1998). Learning from text: Matching readers and texts by latent semantic analysis. *Discourse Processes*, 25(2-3), 309-336.

Athabasca University 

