INNOVATION PROJECT TO VALIDATE AND SELECT ITEMS FOR ASSESSING TRANSVERSAL COMPETENCIES IN HIGHER EDUCATION

J.M. Prats-Montalbán, F. Alarcón, M.M.E. Alemany, A. Boza, M.L. Gordo, M. Fernández-Diego, L. Ruiz, L. Cuenca

Universitat Politècnica de València (SPAIN)

Abstract

The educational improvement innovation project is focused on determining how to evaluate any competence and which are their main related items that should be used for that purpose. Currently, the selection of items is usually performed by groups of experts. However, two main problems related to this type of selection arise in this case: on the one hand, the selection resulting from different groups of experts may be not the same or similar enough, since it is based on the experience and knowledge of each member. On the other hand, the coefficients or weights *a priori* assigned to each item on any competence invalidate any *a posteriori* analysis on its statistical significance and the "real" weight on this competence. To mitigate the above drawbacks, this work presents a methodology able to select, from an objective point of view, the items related to a specific competence, from a set of potentially related ones; furthermore the weights associated to the items are determined. This is carried out by applying a multivariate statistical projection method such as Partial Least Squares (PLS), embedded in a cross-validation process. The paper presents how to preprocess the data, analyze it and obtain the items and their weights to be used for the evaluation of a specific competence.

Keywords: Competence, Items, Variable selection, PLS.

1 INTRODUCTION

Nowadays, there is an increasing interest in evaluating competences achieved by the students by means of a specific or/and a general task, such as teamwork, emotional intelligence, public speaking, etc. This interest has lead to the use of appropriate evaluation tools. Rubrics are one example of such tools that help in the competence evaluation process by testing a group of items. However, the number and which items to use for each competence to be evaluated is not a trivial problem.

The innovation and educational improvement project is focused in determining how to evaluate any competence and which are their principal related items that should be used for it. This is a key aspect for achieving two main goals: 1) to homogenize the items to be evaluated for any competence of interest; and 2) to provide a more objective evaluation methodology.

Currently, the selection of items is usually performed by groups of experts that select, from a wide range of potential items, those having a high degree of relationship with the competence of interest; afterwards they establish a scale for computing a final score on the competence of interest.

Nevertheless, this type of group-of-experts based determination presents some problems. On the one hand, it is not possible to warrantee a unique items selection, since it is based on the knowledge and experience of each group of experts. Thus, different groups may provide, in general, a different selection. On the other hand, a priori establishing a scale, *i.e.* a weight or coefficient for each item, makes not possible to statistically analyze its relationship with the competence of interest. The rationale behind is that the final score in the competence is necessarily affected by the weight or coefficient multiplying each item (*i.e.* it is a linear combination of the latter).

Thus, it becomes mandatory to establish an objective methodology for the evaluation and selection of items, based on statistical evidence. Such methodology starts with a proper data collection, for each and every item included in the original list of potential ones. Afterwards, the competence must be evaluated globally and independently. Only this way it is possible to assess the statistical significance of the relation between each item and the competence. Once all the data have been properly collected, it is necessary to use multivariate statistical models, such as PLS [1] able not only to create

ISBN: 978-84-617-5895-1

prediction models (for the competence from the items) but also to statistically evaluate the degree of relationship with the competence and between the items.

Moreover, by using variable selection techniques (*e.g*. LASSO, VIP's, 2-CV [2-4]) these models let, to determine which are the items presenting a clear statistical significance (i.e. its relationship with the competence is not random); furthermore providing, after the selection, an objective scale that allows predicting a score on the competence from a reduced number of items; which is the final goal of the project.

2 CONTEXT

UPV degrees have a structure based on modules and general subjects. General subjects include the list of subjects and their corresponding number of credits. UPV Transversal skills (CT) aim to synthesize the competence profile acquired by the UPV students, at the same time coping with some degrees with regulations or specific recommendations. A total of 13 transversal competences have been defined. CT1-Understanding and integration; CT2-Application of practical thinking; CT3-Analysis and problem solving; CT4-Innovation, creativity and entrepreneurship; CT5-Design and project; CT6-Teamwork and leadership; CT7-Ethics and professional responsibility; CT8-Effective communication; CT9-Critical thinking; CT10-Knowledge of contemporary problems; CT11-Continuous learning; CT12-Planning and time management; CT13-Specific instrumental (Portal Competencias Transversales UPV, 2016).

Annex I of RD 1393/2007 states that design of each of the subjects must be included, specifying the acquisition of skills and qualifications evaluation system. The new curricula should reflect evidence about the acquisition of the competences gathered in the official titles reports. Therefore, it is mandatory to indicate and properly evaluate these competences.

Degrees are now under certification processes, which must guarantee that they accomplish with the international standards. For students, it is an additional proof recognized in the UE. For employers, it certifies that candidates provide the knowledge and competences needed. For universities, it is an additional proof of the quality and specifies that the certified program fulfills the academic and professional standards.

However this is sometimes a difficult task, since according to the UPV framework about the definition and evaluation of competences, we have now extensive lists of competences, coming from multiple sources and with different approaches in their definitions. Moreover, there are no well-established evaluation mechanisms that ensure the competence acquisition.

These are the issues where the problem is focused and where this project proposal is framed, contributing to the quality of the degrees and guaranteeing the validity.

Related to validation, Messick [5] defines six elements to take into account in order to build a validation frame for the evaluation: content validity (the relevance of the skills and tested knowledge), the external validity (how well the rubrics correlate with other related measures), structural validity (consistency of the internal structure of the evaluation with the domain structure to assess), and generalization (generalization of scoring and interpretations properties across populations).

Ideally an evaluation (for one competence at hand) should be independent of the evaluator and the results should be similar regardless of when and where it takes place. The more consistent it is, the most reliable evaluation (assessment) in different contexts. Of course, different competences may have different weights for the items used, if they are used for the evaluation of more than one competence.

Basically the validity tries to answer the question: Evaluation really measures what you want to measure? The answer to this question is rarely simple. There are two ways to address validation. On the one hand, on the characteristics of the evaluation method itself; and on the other hand, on the interpretation of the results [6, 7].

In the present context, validation is performed on the evaluation method itself, answering to the question: items related to the competence have been chosen objectively and accurately? In this work, we will focus on competence CT4-Innovation, creativity and entrepreneurship. Fig. 1 shows the adequacy, according to the strategic plan of the UPV, of the innovation projects developed by the authors and the new project (indicated by dashed line):

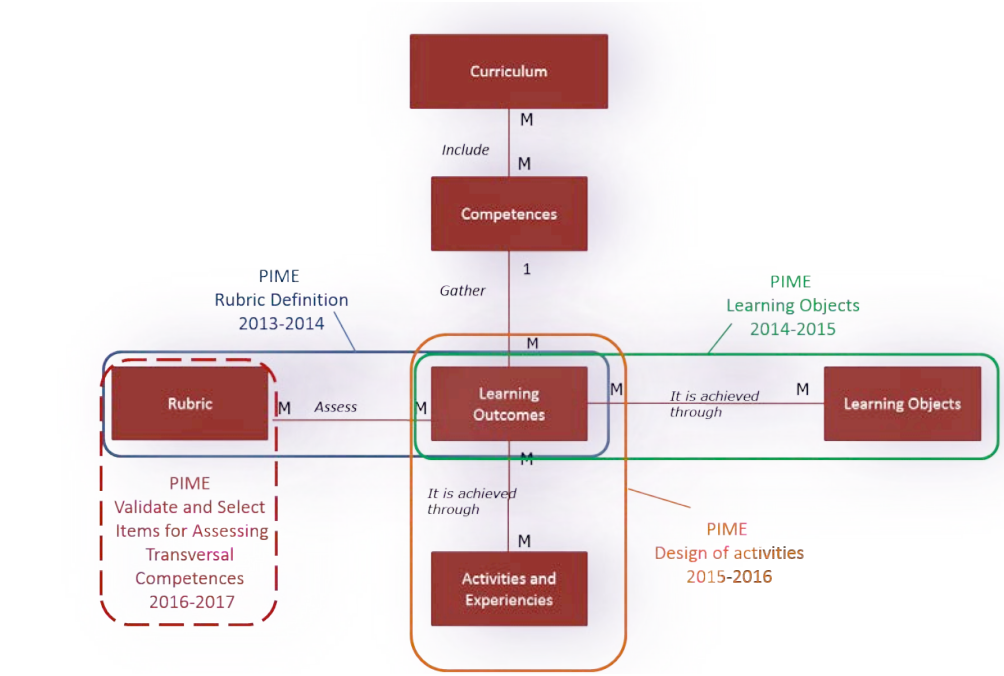

Figure 1: Synthesis of the projects conducted around the development and evaluation of the CT-04.

2.1 Background and analysis of the current situation

To meet the project objective "items associated with the rubric for the competence of creativity, innovation and entrepreneurship have been objectively and accurately chosen", not only the procedure followed, but also the weak points to be improved, should be analyzed. Fig. 2 outlines the process followed in the work of this competence as well as the description of the corresponding steps:

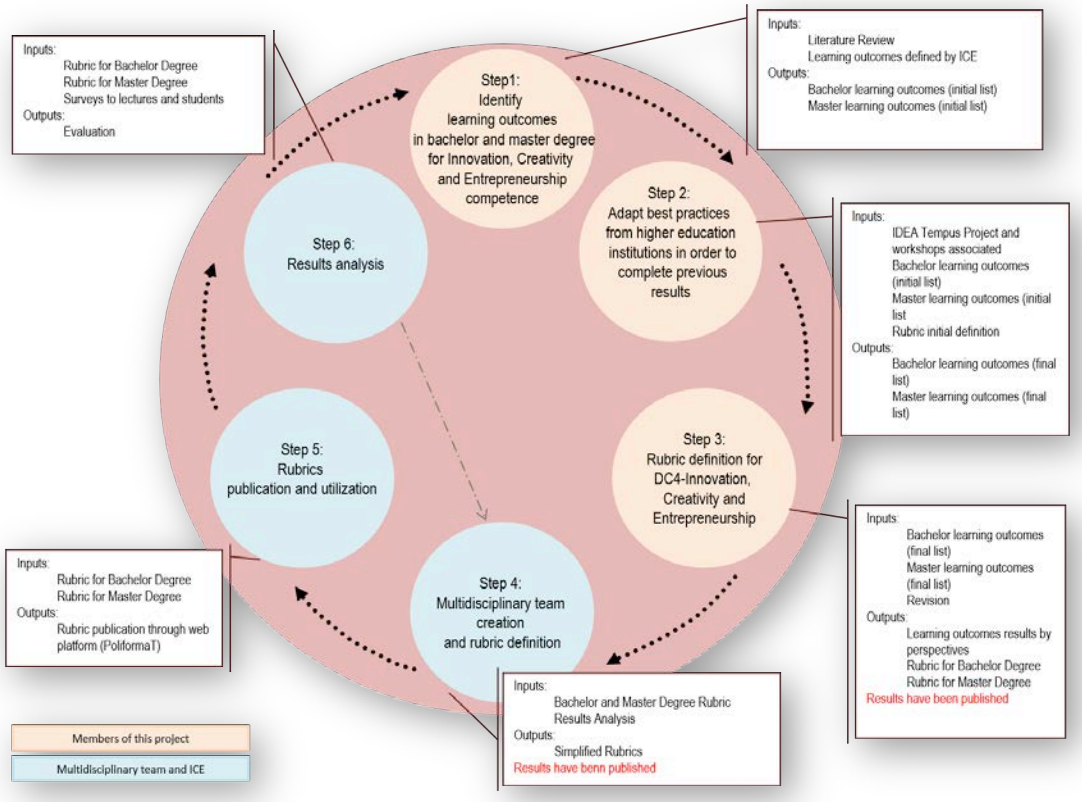

Figure 2: Workflow for CT-04 development.

STEP 1, 2 and 3:

These three steps were carried out by the authors in 2014 under the educational innovation RECICRE project. Developed steps can be considered correct. Step 3 provided as a result the proposed rubric for the Creativity, innovation and entrepreneurship competence, for both degree and master.

STEP 4:

Under the institutional project of the UPV, aware of the need to define a generic section to facilitate the evaluation of skills to all teachers of the UPV, multidisciplinary work teams were created joint to members of the Institute of Educational Sciences (ICE), in order to develop assessment rubrics for each of the thirteen competences.

The multidisciplinary team designed to develop evaluation instruments for this competence (CT-04) was composed of three members of this project requested equipment as well as by other teachers linked to educational innovation teams, technicians from the Ideas UPV Program and educational advisers at the Institute of Educational Sciences (ICE).

From the rubric proposed by our innovation team (result of step 3), we worked on the development of a new simplified rubric (2015). The product of this work was published in the UPV's PoliformaT virtual platform, as a resource to support training and evaluation of CT.

The process followed for the development of the current institutional rubric was based on the literature and on the experience of the "group of experts" in this area. This means that no statistical method was applied.

Step 4 is considered a weak point in the procedure, making it the main goal of this project: to establish an objective procedure, based on the scientific method, for the selection of items (learning outcomes).

STEP 5:

Step 5 corresponds to the dissemination of the rubric defined and its use by teachers. It is currently in progress (2016). This step will be updated once this project with the publication of the new VALIDATED rubric.

STEP 6:

Step 6 corresponds to the analysis of the results, implementation of skills, acceptance and use of rubrics (currently in progress, 2016).

The review of the current situation, the responses of teachers and the applicant's experience project team reinforce the suitability of the application to carry out this innovative project (as it still has not validated any institutional headings generated by these teams UPV) and its extrapolation to the entire university community.

In addition to the innovative nature, the project also stands out for its impact on learning, due to the high number of participating students, as well as the variety of degrees and subjects involved.

The analysis of the current situation has been carried out rigorously, providing empirical evidence for justification and review of the literature related to the purpose of the project.

3 AIM OF THE PROJECT

The main objective of the project is to establish an objective procedure and based on the scientific method for the selection of those items associated with the Creativity, innovation and entrepreneurship competence. In order to deal with this task, it is necessary to define a methodology for evaluating the rubrics related to a specific competence. This way, the methodology could be generalized to any other transversal competence. By applying multivariate statistical models, it is possible to validate and select the statistically significant items, as well as their corresponding coefficients, depending on their weight to the competence of interest.

The rationale for this project is based on the need for objective, easy to understand and to use tools, and for homogenizing the different evaluators criteria when assessing the degree of achievement of some particular competence.

This project continues the work initiated in the RECICRE project (Institutional Project of the UPV) for the incorporation of transversal competences in the curriculum of students, as well as the activities

developed by multidisciplinary teams consisting of teachers and members of the science institute education (ICE).

The project will be implemented and validated on subjects from Industrial Engineering and Computer Science. The defined activities may be transferable to other subjects and other centers of the UPV. The project continues the line of collaboration established in the IDEA-TEMPUS project as well as the new line of collaboration initiated with the Technological Educational Institute of Athens.

3.1 Specific goals

To achieve the overall objective of the project is necessary to carry out several activities and obtain the following results:

1 Define the items used in the study.

   Through review of the literature and previous work done by the innovation team, including meetings between its members, the battery potential items is determined from which it is intended to carry out the selection and subsequent building of the prediction models.

2 Evaluate the competence.

   Evaluate the competence score obtained by each student independently to the evaluation of each item. The evaluation of the competence will be held by the teacher prior to the evaluation of each item, based on the student's global performance during the course. This is necessary because, as mentioned in the project description, evaluation of competences is done many times after the evaluation of the associated items. This procedure misleads any further study on the influence of the items used for calculating the competence, since these items are then often used for the scoring.

3 Evaluate the level accomplished for each item.

   After determining the degree for the competence, the score for each item included in the study is evaluated, in order to perform their validation and selection, by means of multivariate statistical models.

4 Select the items of interest and build the prediction model.

   Once the items have been evaluated, they go through the evaluation and selection process. These techniques allow evaluating the statistical significance of the items in different ways, *e.g.* creating null distributions and calculating the statistical significance of the obtained value for each parameter for each item; or the creation of confidence intervals for the values of these parameters in cross-validation processes. Thus, the statistical significance of the items is evaluated in relation to the competence, bringing them together in a new model to calculate the final global score; furthermore establishing their corresponding weight with regards to the competence.

5 Results

   Finally, the obtained results are to be published in Education and Statistics journals.

3.2 Working plan and methodology

Each of these specific objectives will be implemented through one or more activities to be developed by the project researchers. There will be a person in charge of research activity. The main researcher of the project, besides performing the corresponding own activities, will be responsible for

- accepting the completion of all activities,
- coordinating project tasks,
- monitoring the technical and temporal issues of project tasks,
- ensuring the compliance with the objectives,
- ensuring that the project benefits are achieved,
- organizing the document archive,
- writing down meetings reports,
- being in touch with the managers of the various agencies (ERT, ICE), as well as

- planning and coordinating project dissemination activities (publications , conferences, etc.).

The project-monitoring plan will be carried out through the project team meetings, at least monthly, or whenever the project requires. The Dropbox platform will be used for communication and file-exchange between team members. The meetings will be planned through Google Calendar and Trello application. The meeting place will be the graduate classroom located on the first floor of the Department of Management (bldg. 7D). The project team consists of the main researcher and the different activities leaders. It is a multidisciplinary team of professors from the School of Computer Science, and by professors from the School of Industrial Engineering, with extensive experience in undergraduate and graduate courses and different profiles; allowing synergies between different disciplines such as engineering and business.

3.2.1 Project actions

Action 0: Initial Meeting.

- **Goal**: To implement the activities of implementation of the project.
- **Description**: Meeting of the members of the project team, to present the plan and to formally start the project. Definition of the working environment for the project, the documentary support and workflow to follow.
- **Result**: Report of the meeting.
- **Type**: Document.
- **Leader**: Andrés Boza.

Action 1: Identifying the current activities as well as the potential ones for the evaluation, and the associated learning outcomes for the competence CT4 - Innovation, creativity and entrepreneurship.

- **Goal**: To review and analyze the activities or experiences to be included in the validation and selection process.
- **Description**: review of the documentary sources and of the experience of the team members. This will be the starting point for new designs, and will facilitate a better methodological consistency in the collection and processing of the information.
- **Result**: Document.
- **Type**: Document.
- **Leader**: Llanos Cuenca and Leonor Ruiz.

Action 2: Determination of data collection methodology

- **Goal**: To ensure that all item are measured for their evaluation, in the different subjects evaluated.
- **Description**: Through meetings of the working group, the sample size used, how the different items and the competence will be evaluated; as well as the checkpoints in time (number of evaluations to be performed) are determined.
- **Result**: Database for analysis.
- **Type**: Excel spreadsheet.
- **Leaders**: Faustino Alarcón and Jose Manuel Prats-Montalbán.

Action 3: Exploratory Data Analysis.

- **Goal**: To detect possible anomalous data, as well as groups of individuals and variables that prevent for building local models (i.e. models individualized for some specific subject); or indicate the possibility of creating a single global model.
- **Description**: Once the database has been created, multivariate models (e.g. PCA) that allow studying the internal correlation structure in the data will be used. This lets to detect anomalies in the observations (students), or extreme distances within the model that might cause the creation of "false" latent structures; as long distances to the model, indicative of an internal

correlation structure different from that of the constructed model, hence invalidating any further prediction.

- **Result**: Determination of the number of models to use and report.
- **Type**: Text document.
- **Leader**: José Manuel Prats-Montalbán.

Action 4: Validation and selection of indicators.

- **Goal**: To determine, for each of the possible study populations (local models), which items are statistically significant in predicting the level of competition are acquired.
- **Description**: Apply PLS models (Partial Least Squares) for identifying those items that best predict the value of the competence, by using variable selection techniques such as LASSO (Least Absolute Shrinkage and selection Operator) or cross-validation of parameters related to the items statistical significance (*e.g.* VIP, weights, selectivity ratio). In this way, the best predictive PLS model/s with the most relevant items will be selected.
- **Result**: Items for each subject and/or center, or even globally.
- **Type**: Document.
- **Leader**: José Manuel Prats-Montalbán.

Action 5: Creation of the final model/s based on the corresponding selected indicators.

- **Goal**: To determine, once the items of interest in each case have been selected, the coefficients (scale) that should be used to predict the value of the final competence, for each of the models built.
- **Description**: The PLS models provide, as part of their results, the regression coefficients associated to the items. Therefore, by creating them we obtain these coefficients immediately.
- **Result**: Rubric for each subject and/or center, or general rubric.
- **Type**: Document.
- **Leader**: José Manuel Prats-Montalbán.

Action 6: Validation of the models created.

- **Goal**: To test the ability of the models built to predict the value of competence acquired by new students.
- **Description**: Once the statistical models from the selected items have been created, new data will be obtained using the procedure defined in data collection, and their accuracy on the calculation of the competence will be checked.
- **Result**: Data collected.
- **Type**: Document.
- **Leader**: Mª Luz Gordo and Marta Fernandez-Diego.

Action 7: Review and closing.

- **Goal**: To determine the actions to take, evaluate the possibility of implementing the methodology proposed at a general level for all competitions.
- **Description**: If the methodology is correct, meetings with ICE will be carried out, and the possibility of exporting the results to the whole UPV will be evaluated.
- **Result**: Generalization of the methodology.
- **Type**: Document.
- **Leader**: María del Mar Alemany.

4 PROJECT EVALUATION

The achievements and evidences are reflected in the results. A summary of them may be:

- Creating a complete database with all potential items to be included in the model/s definitive.
- Selection of relevant items in each case.
- Creating predictive models and obtaining the scale.
- Validating the model with new data to verify their predictive ability.
- Implementing new items and the corresponding coefficients for the competence of interest.

4.1 Procedures/instruments for data collection

The data collection instruments are the evaluations conducted by teachers included in this project. This evaluation will be conducted through the generated rubric, with all potential indicators included in the study.

4.2 Data analysis

As indicated throughout the document, the project proposes the creation of a methodology for collecting and analyzing data. This methodology is based on the application of the scientific method, establishing hypotheses and testing them through statistical evidence provided by the data. The treatment of these data will be done through various pre-processing techniques, depending on the type of potential blocks of information that may exist. Their analysis will be done using multivariate statistical projection methods such as PCA and PLS, which analyze the of internal correlation structure of the data, taking advantage of it; as well as using advanced variable selection techniques.

4.3 Potential products

Potential products to be obtained from this project include:

- A new objective and reproducible rubric for evaluating any competence, which can be implemented globally or adapted to any specific subject and/or center.
- A new methodology for determining those items to be evaluated for each competence, as well as the specific scale in each case. This can be generalized for the evaluation of all indicators of all competences.

REFERENCES

[1] Geladi, P.; Kowalski, B.R. (1986) “Partial Least-Squares Regression: A Tutorial”, *Analytica Chimica Acta,* vol. 185, pp. 1-17.

[2] Rasmussen, M.A.; Bro, R. (2012) “A tutorial on the Lasso approach to sparse modeling” *Chemometrics and Intelligent Laboratory Systems*, vol 119, pp. 21-31.

[3] Quintás, G.; Portillo, N.; García-Cañaveras, J.C.; Lahoz, A. (2012) “Chemometric approaches to improve PLSDA model outcome for predicting human non-alcoholic fatty liver disease using UPLC-MS as a metabolic profiling tool”, *Metabolomics*, vol. 8, pp. 86-98.

[4] Kvalheim,O.M.; Arneberg, R.; Bleie, O.; Rajalahti, T.; Smilde, A.K., Westerhuis, J.A. (2014) “Variable importance in latent variable regression models”, *Journal of Chemometrics*, vol. 28, pp. 615-622.

[5] Messick, S. (1996). Validity of performance assessments. In G. Phillips (Ed.), Technical issues in large-scale performance assessment, pp. 1–18.

[6] Borsboom, D.; Mellenbergh, G. J.; van Heerden, J. (2004). The concept of validity. *Psychological Review*, vol. 111, pp. 1061–1071.

[7] Brown, G.; Bull, J.; Pendlebury, M. (1997). Assessing student learning in higher education. London: Routledge.