Journal of the Scholarship of Teaching and Learning, Vol. 15, No. 1, February 2015, pp. 10 - 19.
doi: 10.14434/josotl.v15i1.12905

Collaborative learning supported by rubrics improves critical thinking

Carlos Saiz[1], Silvia F. Rivas and Sonia Olivares

Abstract: In previous works we developed and assessed a teaching program, ARDESOS v.1, with which we aimed to improve the fundamental skills of critical thinking. The results obtained were positive, but modest. After analyzing the limitations of the program we introduced certain modifications and assessed the new version. The changes involved designing the activities programmed by means of rubrics and making the students perform them with less direct orientation from the instructor. In sum specificity and initiative proved to be the key variables in the improved program, ARDESOS v.2. Based on the data collected we have seen a significant improvement of the new version over the old one in the following aspects: a) version 2 improved all the fundamental dimensions, mainly in the pre- and post-test measurements, to a significant extent (Student's t test); b) the effect size (Cohen's d) was significantly higher, and finally c) these improvements in the program elicited better performance. Accordingly, an improvement in critical thinking can be achieved via an instruction design that addresses the factors that really induce change. Currently, with these results we have been successful in adding a new improvement to the instruction, which we have re-evaluated.

Keywords: Critical thinking, Instruction, Evaluation

Introduction

In two previous works (Olivares, Saiz & Rivas, 2013; Saiz & Rivas, 2011) we developed and assessed a program for the instruction of critical thinking (ARDESOS, first version- v.1). The successful functioning of this teaching methodology prompted us to develop a second version of the program (v.2), and also to improve the efficiency of the former version. In the two previous studies, the data obtained were reasonably satisfactory since they reflected important changes in many of the basic skills of critical thinking. This stimulated us to continue working on this ambitious teaching project. The changes observed were also challenging because there were some aspects of the program that did not lead to the expected changes. This is of course quite usual in any line of research: the presence of clearer and more shadowy areas, which should be strengthened and eliminated respectively.

Improving Critical Thinking has and continues to be the underpinning of our research efforts. In our earlier work, we followed several principles and used teaching resources that we have maintained in the present project, although complemented by others. In the first version of the program we used a) the importance of team work, b) direct teaching, c) the need to learn from deficiencies or limitations, and d) the advantages of learning based on problems that arise in people's everyday lives.

Currently, the teaching system has evolved with respect to the first version. A scheme could serve to clarify this. Figure one summarizes the essential features of the ARDESOS v.2 program as used in the present work. In this scheme we have integrated the working methods, tasks, materials and motivational factors. However, to all this we should add, and emphasize, the fact that the participants in the program must decide whether they wish to enroll or not.

[1] Universidad de Salamanca (Spain), csaiz@usal.es. Web: www.pensamiento-critico.com

The students have two options: our instruction program or another conventional teaching program and they must decide which to choose. This choice is more important than it may appear to be. In our program the learning process is based on ideas developed in previous contributions. For the present study, it is appropriate to underscore learning from limitations and problem-based learning (PBL) as the main motor driving the change or improvement in critical thinking. Figure one contains some ideas that are in bold and others that are not. The words in bold differentiate our program from others. They are procedures that have not been implemented or have been used only sporadically. For example, unlike the generalized use of comprehension tasks it is very uncommon to use production tasks in teaching. It is common to use one or another task separately, but not together and the same importance must be given to both. This is one of the original characteristics of our program, at least as far as we know. Moreover, the instruction system based on deficiencies or limitations is certainly one of the most singular aspects of our methodology. Regarding the materials used, there are no studies that have used daily or professional problems, videos, opinion-oriented articles, working the fundamental skills of critical thinking in an integrated way in each of them. As indicated in figure one, these aspects affect and foster the essential motivational traits such as interest, utility, achievement and effort.

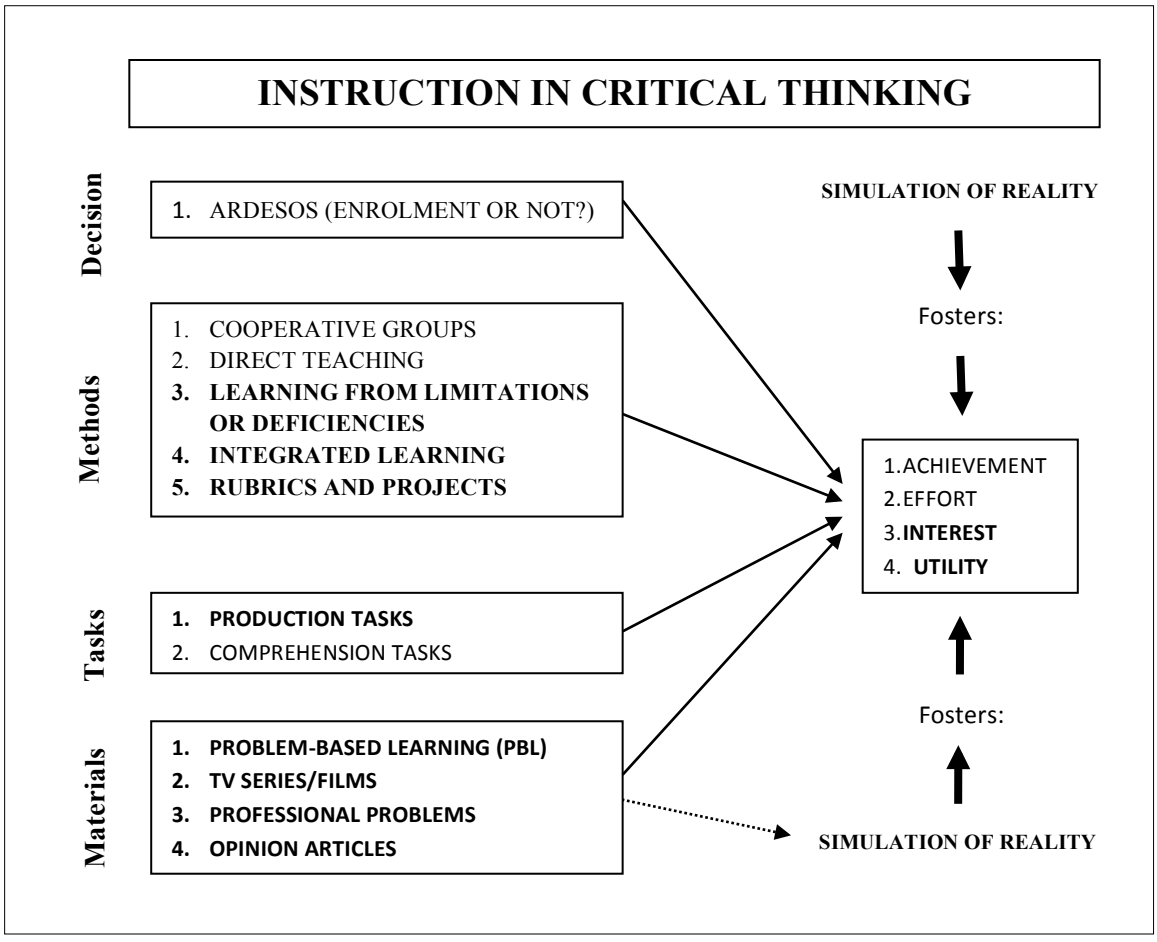

Figure 1. Main characteristics of the ARDESOS v.2. Program.

In the first version of our program, and to a certain extent in the second, the main effort was directed towards achieving efficiency in teaching. It was therefore also mainly directed at achieving an improvement in critical thinking, using strategies, tasks and materials that would guarantee a good result. This global effort to construct a system that would work

was successful. However, we did not know which aspects of the program or which factors or variables were relevant or more relevant than the others. The crux of the matter in the present work is to determine whether there are aspects of the instruction that are more determinant than others. We believe that this is indeed the case. Here we isolated the factors involved in teaching: which of them really makes it work. Also, we wished to know whether it is the overall intervention that fosters the changes in the critical thinking skills. Apparently, this problem in education has not yet been addressed in the literature, but should have been tackled a long time ago. Knowing whether there are relevant factors in instruction is of great importance, both theoretically and practically. Furthermore, in our case, after many years of experience we have observed, but not confirmed, that there are some aspects of teaching that have a greater influence than others on the learning process. One of them has to do with the generalized assertion within the field of education that learning often depends more on what the student does than what the instructor does (Almeida, 2013). The *active participation*, in contrast to the passivity, of students seems to be an especially important factor in education. Nevertheless, there are no studies that have endeavored to check this. Here, we attempt to fill this gap; greater involvement or participation in the learning process must be guided or oriented. Accordingly, active participation by students must be accompanied by *specific instructions*. In the current version of our program the two main changes made are: greater activity or participation in tasks by students, and specificity in the performance of such tasks. How did we operationalize these variables? First by ensuring that the instructor would dedicate more time in directing and orienting the students' work and less time in solving the problems posed. Second, through the elaboration of specific rubrics for each of the tasks or problems posed. This teaching resource made students address the problems by following the indications specified in the method. Accordingly, their activity in the classroom would be focused and well oriented. They knew which aspects were to be worked, the relevance of each, the points they would earn, and the strategies required to apply them.

Thus, the two chief goals in the present work are: a) to determine whether greater activity or participation by the group in resolving the problems posed improves their thinking competencies, and b) whether a guide in the form of rubrics for performing the tasks also contributed to the improvement. Let us illustrate these factors with one of the tasks. One of the activities designed in our program addresses the development of competencies in argumentation. Chart one show one of the rubrics used. It may be seen that the parts to be taken into account and the aspects to be considered in any type of argumentation are detailed and assessed. The method employing rubrics is one of the most efficient ways of quantifying qualitative tasks and guiding learning in a highly concrete and specific manner.

Now, how can we test whether these factors produce change? We tested this by comparing the effect size in the test on critical thinking. The way chosen by us was to compare the effect size in the test on Critical Thinking PENCRISAL (Rivas & Saiz, 2012; Saiz & Rivas, 2008) with the assessment of the ARDESOS v.1 program and the current version. If, as we assumed, the factors introduced in our instruction program determined the improvement in the learning procedure, we expected that the effect size would be significantly greater in the current version than in the previous one. We also expected there would be significant improvements in some of the dimensions of critical thinking that we did not manage to achieve with the first version. Finally, on comparing both programs we expected that performance would be significantly better in version 2 of the Ardesus program. All these changes are addressed in the section on methods.

CRITERIA	SCORE				TOTAL
Comprehension	+5	+5	+5		+15
	Precision in the drafting of ideas	Identification of what is fundamental	Relevant observations		
Argumentation					
	10	45	+10	+5	55/+15
Structure	Conclusion	5 mainreasons/ counterarguments	Another 3 reasons/ counterarguments	Restrictions or conditions	
	+5	+5	25	+5	25/+15
	Opinions, assumptions, conjectures,...	Facts	Relations	Other Considerations	
	5	10	5	+5	20/+5
Assessment	Acceptability	Relevance	Global	Falacies	
MAXIMUM TOTAL SCORE	15/+10	55/+10	30/+15	0/+20	100

Chart 1. Rubric Arg.1 Group Comprehension Task Argumentation.

Methodology

Participants

The sample of the present study comprised 144 students from the first year of the Degree in Psychology of the University of Salamanca. Of these, 82.6% (119) were women as compared with 25 men (17.4%). This difference is statistically significant (Chi^2= 16.531; 1 gl; p=.00). The mean age of the sample was 18.83 (s.d. 1.89) (CIat 95%: 18.51-19.14), within a range of 18-32 years. The distribution did not fit the normal model with p<.01 on the Kolmogorov-Smirnov goodness of fit test (p=.00) owing to a marked positive asymmetry (As= 4.00) and a clearly leptokurtic shape (K=20.40). The study sample of version v.1 is described in the paper by Saiz and Rivas (2011).

Instruments

The PENCRISAL Critical Thinking test.

This test comprises 35 situation-production problems, with an open format and is structured around 5 factors: Practical Reasoning, Deduction, Induction, Decision Making and Problem Solving (Cronbach alpha = .632; test-retest = r =.786, Rivas & Saiz, 2012). Each of the factors contains the most representative structures, thus enabling us to isolate the main skills of Critical Thinking and the most relevant methods of reflection and resolution of our daily lives. The PENCRISAL test has been described in detail in Saiz and Rivas (2008). This test was designed following the methodology of task analysis in order to uncover which processes or mechanisms of thinking are functioning on each of the 35 problems posed in the test. The problems were designed in such a way that it was only posible to solve them by using a strategy or a mechanism. Thus, we know that on solving a problem, an item of causality, this can only be done using causal reasoning and not other mechanism. In other words, that if a problem needs to be solved using an identification strategy it cannot be done in any other way. What is more important, we can identify the mechanism in the open answers given on the test. For further information, the links to those works in English can be consulted:

Saiz and Rivas (2008):
http://www.pensamiento-critico.com/archivos/evaluationCTergoENGLSH.pdf
Rivas and Saiz (2012):
http://www.pensamiento-critico.com/archivos/validacionpencrieng.pdf

ARDESOS v.2 Program

As reported above, in comparison with the first version the instruction was improved. The duration of the program was 60 hours (face-to-face teaching) along 15 weeks and four hours of class per week. The instruction was given in classes of 30-38 students divided into four groups so that the students could work in teams. All activities were planned at the beginning of the course, with rubrics. The classroom work was directed towards the development of these activities, under the supervision of an instructor. The role of this latter consisted of orienting the students in each of the tasks and clarifying any doubts that might arise during their completion. Later, in the assessment of the activities the solution to each activity was explained to the students.

Assessment was performed on a weekly basis, with feedback facilitated 2 to 3 days later. The importance of the immediacy of the assessment should be noted in the sense of that it fosters a good development of learning. The assessment was quantitative, as specified in each rubric. Thus, students knew how much weight each part of the task carried and what was more or less important. For example, in an activity involving argumentation what was most important were the identification and relationships of the elements of an argument, while its evaluation was less important. It is important to recall that evaluation is an essential component of our program; the learning process would be impossible without it.

Procedure

The ARDESOS v.2 program was applied along one term at the School of Psychology of the University of Salamanca. One week before the start of instruction all students took the PENCRISAL test. Likewise, it was applied one week after the intervention to obtain a second measurement of the variables. The time elapsed between the pre-treatment and the post-treatment measurements was four months. The first version of the ARDESOS program was implemented using a procedure identical to that used in the application of the current one.

Design

In order to analyze the efficiency of the intervention we used a quasi-experimental design, with pre- and post-treatment measurements.

Statistical analysis

In the statistical analysis we employed the IBM SPSS Statistics 19 package. The tools and statistical techniques used were as follows: frequency tables and percentages for the qualitative variables, with a Chi-square test for homogeneity; exploratory and descriptive analyses of the quantitative variables with a test for goodness of fit to the normal Gaussian model and box diagrams for the detection of atypical values (outliers); statistical techniques (mean, standard deviation, median… etc.) for numerical variables; the *t* test for the value of a measurement, tests of the significance of differences of Student's *t* means, and calculation of Cohen's *d* to estimate effect size.

Results

Regarding the descriptive statistics of all the variables included in the study, we observed that most of them fit the model of normality adequately, although some had significant deviations, which were overlooked due to the size of the sample.

Table 1

Descriptive statistics of the variables

	N	M	SD	Minimum	Maximum
TOT_PencriPre	144	28.58	6.53	12	45
DR_PencriPre	144	3.98	2.00	0	9
IR_PencriPre	144	5.06	1.81	1	16
PR_PencriPre	144	6.31	2.53	0	12
DM_PencriPre	144	6.69	1.94	1	11
PS_PencriPre	144	6.53	2.19	1	10
TOT_PencriPost	144	31.70	6.49	14	44
DR_PencriPost	144	5.25	2.17	0	11
IR PencriPost	144	5.48	1.67	2	9
PR_PencriPost	144	8.40	2.32	1	13
DM PencriPost	144	7.01	2.08	2	13
PS_PencriPost	144	5.56	2.49	0	11

Below, the results of the statistical analyses performed are shown in order as a function of the above aims.

In order to assess the differential effect caused by the program over the two years and to determine in which factors the improvements introduced were affected the most, we performed tests on the significance of differences of Student's *t* means and calculated Cohen's *d* values to estimate the effect size.

As can be seen in table two, the results provided by the descriptive statistics indicate that the optimized v.2 ARDESOS program was more effective since it significantly improved the performance on the post- measurements across the whole scale and in all the factors, with the exception of decision making, whereas with v.1 a significant increase occurred only in the post-performance of the induction and decision-making factors.

With a view to analyzing the impact of the intervention of the two versions, we used the standard mean difference, *d*, of Cohen (1988) as an index of effect size. The data show that in v.2 of the program a significant increase occurred in the deduction, practical reasoning and problem-solving factors and in the overall score of the scale. It may be seen that regarding practical reasoning (*d*=.83) and deduction (*d*=.63) effect size has very high values. However, in v.1 these values are lower (Pract. Reasoning: *d*=.03; deduction, *d*=.45). Likewise, the total of the scale (*d*=.48) and the problem-solving factor (*d*=.44) had a moderate effect size in v.2 whereas in v.1 these values were very low, ranging around .10. In light of these results, it may be concluded that the improvements introduced are reflected in an increase in critical thinking skills and the skills with the greatest effect size are practical reasoning and deduction, followed by problem-solving and, to a lesser extent, induction skills.

Table 2

Differences in Student's t means and effect size- Cohen's d

	ARDESOS PROGRAM VERSION 1					ARDESOS PROGRAM VERSION 2				
	PRE	POST	Student's *t* test			PRE	POST	Comparison		
	M (SD)	M (SD)	Diff. in means p-sig n	t	Effect size	M (SD)	M (SD)	Diff. in means p-sig n	t	Effect size
DED	6.31 (2.47)	5.21 (2.21)	1.10** .000 97	3.83	.45	3.98 (2.01)	5.25 (2.17)	-1.27** .000 144	-6.57	.63
IND	3.74 (1.59)	4.69 (2.20)	-.95** .000 99	3.84	.60	5.06 (1.81)	5.48 (1.67)	-.41** .006 144	-2.51	.23
PR	6.37 (2.69)	6.47 (2.74)	-.10 .741 97	.33	.03	6.31 (2.53)	8.40 (2.32)	-2.09** .000 144	-9.08	.83
DM	6.08 (1.74)	6.64 (2.04)	-.56* .040 88	2.08	.32	6.60 (1.94)	7.01 (2.08)	-.31 .063 144	-1.53	.16
PS	3.75 (1.32)	3.53 (1.29)	.22 .135 94	1.51	.17	5.56 (2.19)	6.53 (2.49)	-.97** .000 144	4.72	.44
TOT	25.98 (6.27)	26.65 (7.35)	-.67 .448 88	.76	.10	28.58 (6.53)	31.79 (6.49)	-3.12** .000 144	-5. 87	.48

* Significant at 5% ** Significant at 1%

Since we were interested in checking whether the improvements might indicate better performance we decided to use the *t* test to see whether the values of the means were statistically significant (see table 3). We then compared the means of the improved version with the average mean obtained in the v.1 sample. The difference between the means of the improved version and v.1 proved to be statistically significant at $p<.01$ on the whole scale and on all the subfactors, except decision making. This allowed us to conclude that the sample analyzed with the improved version of the program afforded a significantly better performance than the sample of v.1.

We observed that the sample analyzed had a significantly improved performance on all the skills of critical thinking than (as compared with) the sample from v.1, with a difference of 5.5. points (CI 95%: 3.98-6.12). Regarding the critical thinking skills variables, we noted that all of them but one underwent a statistically significant increase. Deduction rose from a mean of 5.21 in the first version to 5.25 in the second one (CI 95%: .31-.40). Although the means are fairly similar, it should be noted that in v.1 there was a problem in the pre-measurement because the instruction had already been followed, such that –as seen in Table 2- it was higher than the post- value. The result of the second version is therefore important since the increase from the pre- mean to the post- mean was more than almost a whole point, accounting for .63 of the effect size. Induction was affected to the same extent, with a significant increase in its mean of almost a whole point (CI 95%: .51-1.06). Practical reasoning also showed higher performance means in the second version, where an increase of almost two points was observed (CI 95%: 1.55-2.31). The decision-making variable evolved in a similar fashion to the others, although the analyses revealed a small increase (.367) in the

mean of v.2 (CI 95%: .02-.71). Finally, problem solving had the strongest increase in its mean (CI 95%: 1.62-2.44). In the first version, the students obtained a mean score of 3.52 whereas in the second version the mean rose to 6.53).

Table 3

Student's t test for the contrast of hypotheses for the value of a mean

Variables	Contrast value for the mean	N	M	SD	Difference (CI 95%)	Student's t test	
						T	P-sign
DED	5.21	144	5.25	2.17	.40	.220	.413
IND	4.69	144	5.48	1.67	.789	5.663	.000**
PR	6.47	144	8.40	2.32	1.933	9.978	.000**
DM	6.64	144	7.01	2.08	.367	2.116	.018*
PS	3.53	144	5.56	2.49	2.033	9.769	.000**
TOT	26.65	144	31.70	6.49	5.505	9.336	.000**

* Significant at 5% ** Significant at 1%

Globally, the results support our predictions since we observed important changes with v.2 of our program. Properly directed, greater participation and more collaborative work mean that the improvement in critical thinking skills is substantially greater. We observed that the only change in instruction, with version 2 of the program, was greater *activity* and *specificity*; all the rest remained equal. Accordingly it would be reasonable to speculate that these variables would be responsible for the results. We go further into this in the Discussion section.

Discussion and conclusions

Having discussed the analyses, we are now in a better position to assess the progress made in our second version of the ARDESOS program (v.2). Previously we stated that we were seeking to determine whether a change in critical thinking had occurred from one type of instruction to the other in three ways: 1) comparing the effect size in the test of critical thinking; 2) observing whether an improvement had been achieved in the dimensions of critical thinking for which satisfactory results were not obtained with the first version (v.1), and 3) observing whether performance was better with the new version of the program. The above analyses show that with the new version of the program the effect size was considerably improved, leading to a change in all the dimensions of critical thinking. However, in decision-making a positive improvement was observed as regards trend but not with respect to significance. Additionally, the very high values obtained for the effect of the practical reasoning and deduction dimensions are promising. We believe that obtaining these values with the changes introduced into the program means that we should be optimistic or expect similar results in the other dimensions. These, argumentation and deduction, are the dimensions best delimited conceptually, and decision making and problem solving are the least well delimited. Accordingly, once greater precision has been achieved in these latter two, we expect to obtain similar results in these four dimensions.

Across the whole scale and in problem solving the values were moderately high. Regarding the improvements with respect to the pre-post differences, we obtained the same pattern of changes, namely an improvement in all dimensions. However, despite the

observation of a positive trend decision-making did not reach statistical significance. Finally, performance in critical thinking improved across the whole scale. This was especially the case of induction, practical reasoning and problem-solving with respect to the first version of the program. Concerning decision making, performance was moderate but acceptable. However, performance in deduction did not improve owing to an anomaly in the procedure used in the first version (see above).

From the foregoing, our conclusions are clear. The results expected from our approach are very positive, with the observation of an effect size, pre-post differences and performance that were quite high across the scale and in some of the dimensions. Only decision making failed to meet our expectations, this dimension showing modest and in some cases non-significant values. By contrast, the problem-solving dimension improved considerably. To understand this lack of consistency in the data -a slight change in decision making and a large change in problem solving- it should be recalled that both dimensions share general items, two and four respectively. The instruction in the current version of the program works the general process of problem solving much more intensely and places less emphasis on specific strategies. A possible explanation for this may lie in the fact that decision making does not benefit from the change in instruction, unlike problem solving. It should also be noted that there is a conceptual difficulty involved in separating these general strategies from these dimensions. The difference between these two dimensions is not clear, because both of them have general items, and it is difficult to know whether they are general items of problem solving or decision making. This is essentially a conceptual problem that we are currently trying to solve.

From the modifications in the instruction corresponding to the current version of the program it may be suggested that in part the problem could be solved by approaching these strategies based on a single factor (efficiency). This means that they would be used in a context of choice or of solution to obtain the best result possible.

The current improvements in our instruction program partly contribute to solving this conceptual problem. One way of solving it is to use strategies guided by a factor common to the general strategies of problem solving and decision making. This factor is efficacy, which will drive all the strategies, in order to obtain the best result possible or the best solution to the problem approached.

Our prognosis is that these conceptual and empirical difficulties will disappear. In fact, we already have one result pointing in this direction, since having the best explanation of a problem guarantees maximum efficacy and with this many action strategies become superfluous. However, will be addressed in a future work.

References

Almeida, L. (2013).*Ajustamento e sucesso acadêmico no Ensino Superior: Um roteiro para a avaliação das suas dimensões psicológicas.* IV Congresso Brasileiro de Avaliação Psicológica. Maceio 4-7 de Junio de 2013.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Olivares, S., Saiz, C. & Rivas, S.F. (2013). Encouragement for thinking critically. *Electronic Journal of Research in Educational Psychology, 11*(2), 367-394. doi: 10.14204/ejrep.30.12168

Rivas, S.F. & Saiz, C. (2012). Validación y propiedades psicométricas de la prueba de pensamiento crítico PENCRISAL [Validation and psychometric properties of the test of critical thinking PENCRISAL]. *Revista Electrónica de Metodología Aplicada, 17*(1), 18-34.

Saiz, C. (2013). *Pensamiento crítico como reto personal y social*. Palestra en el Programade Pos-graduaçao Stricto Senus en Psicología da Universidade Sao Franciso. 13 de Novembro, 2013. Itativa, Sao Paulo (Brasil).

Saiz, C. & Rivas, S.F. (2008). Evaluación del pensamiento crítico: una propuesta para diferenciar formas de pensar [Assessment of critical thinking: a proposal to distinguish ways of thinking]. *Ergo, Nueva Época, 22-23* (marzo-septiembre), 25-66.

Saiz, C. & Rivas, S.F. (2011). Evaluation of the ARDESOS program: an initiative to improve critical thinking skills. *Journal of the Scholarship of Teaching and Learning, 11*(2), 34-51.

Saiz, C. & Rivas, S.F. (2012). Pensamiento crítico y aprendizaje basado en problemas [Critical thinking and problem-based learning]. *Revista de Docencia Universitaria, 10* (3), 325-346.